The Acquisition of Phonology in the First Year of Life

Philip Archibald Harrison

University College London
University of London

1999

Thesis submitted to the University of London
in partial fulfillment of the requirements of the Degree of
Doctor of Philosophy

Abstract

Any phonological theory needs to encompass an account of acquisition and any account of acquisition must take its place within a general theory of phonology. This thesis aims to ascribe phonological significance to speech perception in infancy, a move impossible unless phonology is defined, as it is here, from both a psycholinguistic and a formal viewpoint as a dedicated pattern-recognition system. Extant results from infant studies are reviewed and aligned with current phonological theory. In particular, such theory characterises phonology as bi-modular, so the acquisition of individual melodic and prosodic modules and their subsequent orientation with respect to one another must constitute three different developmental tasks. This delivers a relatively simple account of the mapping between psychoacoustics and phonology.

Perception and pre-existing theories of segmental complexity are related using an original experiment into the perception of vowel-height contrast in Catalan.

If infant perception has phonological import, then disparate phonetic reflexes which are predicted as phonologically identical should show parallels in acquisition. General theory argues that the same abstract melodic objects underlie both laryngeal contrasts in stops and lexical tonal contrasts. Earlier studies show that language-specific attunement to stop contrasts has taken place by the age of six months. New tests are now reported, using children of the same age, which demonstrate that infants acquiring Yorùbá, a language which has a three-way contrast for tone, attend more closely to pitch changes within the minimal domain word than do English controls. Further, they only attend to those pitch changes that possess phonological import within that domain in the steady-state language. In this their perception exactly parallels that displayed by adult speakers. Apparent anomalies in the results of these tests are shown to be closely parallelled by phonological asymmetries in the tonology of Yorùbá.

Table of Contents

Chapter 1
Aims and Definitions

Chapter 2
Phonology and Infants' Perceptual Abilities

Chapter 3
Language-Specific Prototypes in Catalan

Chapter 4
Experiments with Tone

Chapter 5
Yorùbá and Nuclear Laryngeal Elements

Chapter 6
Contexts

Appendices

Illustrations

PAGE
MISSING
IN
ORIGINAL

Acknowledgements

John Harris has not bought all the ideas in this thesis, but he has certainly paid for them. I am enormously grateful to him for using up vast amounts of his valuable time reading and re-reading drafts and pointing out things which hadn't occurred to me with good humour, patience and often terrifying accuracy.

I have also been very lucky in receiving multifarious professional assistance from other members of staff in the department of Phonetics and Linguistics at UCL. Andy Faulker demystified statistics for me, Mark Huckvale took time out to create the software that enabled the perception testing paradigm, Steve Nevard designed and built all the necessary circuitry, and Andrew Simpson acted as my unpaid tutor in speech analysis and synthesis. Outside the department, I am grateful to Mark Leaity for his mathematical expertise.

My prinicipal informants were Akin Oyètádé for the Yorùbá data and Eva Estebas Vilaplana for Catalan: many thanks to both for their indulgence and enthusiasm, and to the other adults who participated in the perception tests. Nothing would have been achieved without the cooperation of the parents of the infant subjects, who were hugely generous with their own time and that of their offspring.

Thanks also to the staff at Wolfson House for allowing some of their valuable space to be used up, and particularly to those staff and students kind enough to assist with the experiments. These experiments were made possible by financial assistance from the department and from the UCL Graduate School.

Lastly, thanks to my wife Francesca, who could make almost anything happen.

PAGE
MISSING
IN
ORIGINAL

Abbreviations and Notational Conventions

C	consonant
ERP	event-related (change in) potential
Fx	fundamental frequency
GP	Government Phonology
HAS	high-amplitude sucking
Hz	Hertz
JND	just-noticeable difference
ms	milliseconds
N	nucleus
NLM	Native-Language-Magnet
NLT	*No Low Tone* (hypothesis)
O	onset
OCP	Obligatory Contour Principle
OT	Optimality Theory
pl.	plural
PrWd	Prosodic Word
R	rhyme
RP	Received Pronunciation
SW	Subword
V	vowel
VOT	voice onset time
VR	visual reinforcer
VRISD	visually-reinforced infant speech discrimination (paradigm)
dB	decibel
dBHL	dB hearing level
dBSPL	dB sound pressure level

Both lexical tones, when abbreviated, and elements are indicated using capital letters. The notational distinction between them is signified by the use of square brackets. Thus H = high tone, while [H] = element [H].

Transcription, which conforms to the IPA system, is no narrower than necessary. In running text it is, as usual, enclosed in square brackets ([]). Slanted brackets (//) are only used if focus is being drawn especially to a lexical contrast in one particular language. Non-English examples represented orthographically are enclosed in angled brackets (< >). Brackets are not generally used to enclose transcription in numbered examples, where square brackets ([]) may be used instead to signify prosodic domain boundaries.

Age is notated thus: yrs:mths:days (e.g 0:7:14).

Chapter 1

Aims and Definitions

1.1 Overview

It so happens that the department of Phonetics and Linguistics at University College London occupies two sites. Experimental Phonetics is separate from Linguistics. To get from one site to the other, it is necessary to cross the Euston Road, a six-lane urban canyon with torrential traffic flows. This has a certain allegorical resonance not lost on sensitive souls, and it inspires bridge-building, which is what this thesis is all about. In the main, the discussion will draw upon research in the fields of theoretical phonology, the typology of languages, and the auditory perceptual abilities of infants, with roughly equal indebtedness to each. These will be tied together with the aim of convincing the reader that human ontogeny in the first year of life includes the development of a language-specific phonology. If this argument is accepted, an inevitable consequence is that informing infant perception tests (and perception and production experiments on older children) with phonological theory opens up new vistas for research.

Setting a discussion under the title 'The Acquisition of Phonology in the First Year of Life' involves a commitment to a relatively radical position. Ingram (1989) 'cautiously' proposes that children may begin to acquire the phonological system of their native language during the developmental period of single-word utterances (Ingram 1989: 218). Here, the argument will be advanced that there is no reason why children should postpone phonological acquisition until this relatively late stage (post one year), but rather that they may utilise their attested sophisticated auditory (psychoacoustic) perception during the first year of life to parametrise universal phonological primes. What may well be postponed is the integration of the prosodic and melodic representational levels. That such levels maintain a degree of independence into the adult language is a crucial proposal of some recent phonological theory, and we will find evidence during our discussion of acquisitional facts that bears directly on this claim.

There is certainly no logical reason to disallow the proposal that the acquisition of a phonological system, or of an autonomous *part* of a phonological system can be chronologically independent of the acquisition of any other modular component of the grammar. The chief focus of the present discussion is the acquisition of melodic (qualitative) contrasts and the adequacy of our theoretical model to deal with this acquisition. Our theory-internal justification for looking at this development in relative isolation inheres in the following two proposals, detailed in Kaye, Lowenstamm & Vergnaud (1985), Harris (1994) and Harris & Lindsey (1995), among others:

(1)

(a) melodic primes can be represented as individual abstracts (elements) and

(b) these are independent of prosodic structure, with which they interact via relationships which may be expressed in terms of licensing.[1]

By adopting this theoretical position, it becomes inevitable that the phrase 'acquisition of phonology' is rendered *formally* meaningless. If the mature system is represented as two interactive subsystems which largely retain independence, 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 does not seem remotely defensible, and would require a general account of how it is possible to devolve two linguistic structural systems from one supersystem.

To exemplify from elsewhere within the language faculty, 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: 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 morphological word boundaries, we deliver up the unlikely hypothesis that <na ten stúl> ('onto that table') is three words, and <ná stul>

---

[1] These two assertions will be expanded upon in §1.4, and in chapters 2 and 6.

('onto the/a table') is one. Bracketing paradoxes in English, wherein morphological and phonological derivations may not proceed in coincident cycles, provide further confirmation of the lack of necessary isomorphism between the two systems.[2] They must somehow interface. Explaining the *nature* of the interface has of course been the subject of a good deal of serious endeavour, but its existence, in some form, is not in dispute.

So it is within the phonological model adopted here. During acquisition, simplex structures *within a single system* may bifurcate or differentiate. The acquisition of branching onsets, and of the representations for palatality and labiality are two examples, one taken from either side of the skeletal divide. Some form of analysis and synthesis of congeneric objects is crucial to other aspects of phonological development (for instance, the acquisition of vocalic primes), but nobody has ever tried to advance the untenable notion of prising apart any two discrete structural *systems* during ontogenesis. If there are two systems in steady-state language, then they must undergo separate development.

We discuss this matter further in the light of extant acquisitional literature in chapter 2, when we will make our first claims that characterising early acquisition as 'phonological' solves some apparent anomalies in previous research. This enables the design of two research programs, one of which concerns the phonological identity of phonetically disparate objects, and occupies centre stage for most of the following chapters.

Chapter 3 reports an experiment concerning the perception of vowel heights by Catalan and English native speakers. This is intended to provide some original confirmation of the existence of 'prototypes' and more particularly to relate these prototypes to a language-specific phonological system. The results may ultimately have some bearing on a putative connection between complexity and chronology in acquisition which naturally falls out of our stated view on infant perception and its relation to phonology. They do at least serve as an example of the potential usefulness of informing empirical tests with theory.

---

[2] Bracketing paradoxes are exemplified in §5.4.2 (p.153), during a discussion of 'phonological word-edge'.

Armed with this reassurance, we turn in chapter 4 to the main focus of the present work. A series of experiments is described involving both adult and infant subjects from Yorùbá and English backgrounds. These experiments have been designed to investigate a second hypothesis that results from the basic tenet that infant perception has language-specific phonological significance, namely that phonetically disparate objects may show acquisitional parallels if they are phonologically identical. The results reported to some extent support this hypothesis, and so in turn the general theory, in conformity to Jakobsonian ideals. Chapter 5 consists of a reappraisal of the tonal phonology of Yorùbá: this is a necessary parallel to the empirical tests if we are to make full sense of the findings reported in chapter 4.

Chapter 6 is an attempt to contextualise the content of the earlier chapters in a number of ways, and so demonstrate that the direct link proposed between the theoretical model of melody espoused here and acquisition makes general sense. Chapter 7 summarises what has gone before, provides some concise answers to the questions that have been raised, and exemplifies how the present approach could be valid for other applications.

1.2 Aims

Setting empirical perception testing firmly within a theoretical linguistic/scientific framework (as is desirable) means that it is possible to focus on several questions at once. In the present case, we may hope on the one hand to answer a simple enquiry about infants' perceptual abilities, but simultaneouly we may aspire to adding a few drops to the ocean of knowledge and speculation about where phonology 'is' in the language-module, and how many representational levels it is optimal to propose as extant on the 'phon-' side of linguistics.[3] The existing answers to these bigger questions are far from achieving universal acclaim, and sometimes the questions themselves languish unasked by researchers. There follows a list of the four main related questions that we hope to get

---

[3] For a review of various previous answers to this question, particularly in the light of the contradictions that may be inherent in those answers, see Burton-Roberts & Carr (1996: 14-27).

partial or whole answers to in the present work.

(2)

(a) Can the lights of phonology illuminate existing acquisitional findings, and do existing acquisitional findings integrate with phonological theory, especially with regard to the prosody/melody bifurcation?

(b) Is it possible to design infant acquisitional tests to answer properly formulated, specifically phonological, questions?

(c) Can we deliver an account of the acquisition of melodic primes?

(d) Do babies display evidence of language-specific *phonological* processing[4], such that

    (i) parallel acquisition of items that are phonetically diverse, yet phonologically identical, may be demonstrated, and

    (ii) there is any traceable connection between phonological complexity and acquisitional chronology?

We will return to these questions innumerable times during the following discussion. It is hoped that trying to find answers to them will contribute a little to the alleviation of the theoretical dearth often complained of in the literature. Aslin & Pisoni (not untypically) remark: 'Although there is currently a great deal of empirical data available on infant speech perception, the same cannot be said for theories of development' (1980b: 94). In the ensuing eighteen years since this was written, not a lot has changed. Kaye (1997) still finds grounds for admonishment: '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: 219).

Having outlined our ambitions, it may be helpful to say something about the limitations

---

[4] Their ability to display language-specific 'phonetic' (or rather 'psychoacoustic': see §1.3.3.below) processing has long been established, as we will see in chapter 2.

we place on them. The empirical testing reported here will be confined in scope to the *exemplification* of informing such tests with phonological priorities. To fully underpin a theory of early language development with experimental results, while highly desirable, is beyond currently available resources, and also ideally awaits the solution to some methodological problems (which are discussed in §.4.4). So while the tests in the main have their genesis in a single theoretical nugget (the proposed phonological link between stop contrasts and lexical tone[5]), this is not exploited because of its status as a *unique* facet of theory, but rather as a potentially typical opportunity for the design of a research programme that may unite perception, acquisition and phonology.

1.3 Definitions

1.3.0 Apparently basic definitional statements about the nature of 'phon-' representational levels must be made at the outset, particularly in a piece of work that seeks to correlate more than one of these 'levels'. Some of these are only *apparently* basic definitions, because consensus on the status of 'phonology' and the like does not exist. Tacitly or otherwise, phoneticians, phonologists, and linguists in general do not agree on where to place the interfaces either between different cognitive structures, or between the physical and the abstract. While it is hoped that the work reported here may help to illuminate this issue a little, it is reponsible first to make a commitment to a jumping-off point, even if it is not one that everybody agrees with. In this spirit, the following definitions are offered. They are qualified with examples where this seems necessary.

In order to avoid as much obfuscation as possible at the outset, adjectives rather than nouns are the subjects of the proffered definitions. This is simply to focus attention on the uses of these terms which are presently relevant. Nouns have become closely associated with competing (or supplementary) theories: thus 'Declarative Phonology' *versus* 'Lexical Phonology' *versus* 'Government Phonology' etc. Though we gratefully exploit the theoretical advances that have been achieved by 'Government Phonology' in the ensuing

---

[5] See §2.4, (p. 57).

chapters of this discussion, our interest at first is in characterising as pre-theoretically as possible the different types of cognitive abilities we wish to identify and correlate. This may enable an understanding of the meaning (or lack of meaning) in questions such as: 'at what point does perception cease to be '*psychoacoustic*' and become '*phonetic*'?

1.3.1 'Acoustic'

Anything 'acoustic' relates to the physical properties of sound, which are entirely modality-neutral and directly quantifiable. For speech, or for music, or for noise, these properties may be durational, frequency-related, intensity-related, or any compound of these three physical parameters measureable off a sound spectrum or temporal representation.

1.3.2 'Psychoacoustic'

Anything 'psychoacoustic' relates to the perception of sound by living organisms. This perception may in some cases be directly correlated with acoustic parameters: thus dBHL (hearing level) uses a 'psychoacoustic'[6] baseline which is ultimately derived from the reactions of a group of organisms (people) to physically measured changes in atmospheric pressure. In other cases, no direct acoustic correlation may be possible: thus the imposition of a categorical boundary in the Voice-Onset-Time continuum[7] may ignore any consideration of acoustic salience. Psychoacoustic perception may or may not be utilised for the ecological advantage gained for an organism by efficient communication: in humans only it may be harnessed in the service of language. Thus in the present case we use the term 'psychoacoustic' to refer to the perception of sounds which are not linguistically significant. This definition is a negative one, but is of use in a discussion of language even though it lumps together a vast and interestingly variegated range

---

[6] The term 'psychophysical' is also used.

[7] See §2.2.1, (p.41).

of communicative activities.[8] It puts paralinguistic phenomena on a par with other unstructured, holistic communications between animals, which seems correct, since it is obvious that animals can certainly 'mean' something. Observation of the behaviour of an animal with a highly developed imitative ability *and* the physiological capacity to create speech-like utterances (e.g. a parrot) in a situation where it is patently using these utterances in direct response to that situation can show quite clearly what is possible without any linguistically dedicated capability, and with a purely holistic 'vocabulary'.

It is worth also noting that sound *structures* are not unique to human linguistic communication. Passerine birds clearly impose structural constraints upon their song productions. In the song of the mistle thrush, for example, 'only five themes appear at the beginning, the same theme never both precedes and follows a given theme, most themes are never repeated,' ...and... 'most themes can be followed by no more than one or two other themes' (McNeill 1970: 45). The ecological value of this ability is entirely mysterious. No communicative purpose has been attributed to these productions: the hierarchy of sound structures does not, apparently, reflect an ability to combine messages. There is equally no suggestion that the thrush's repertoire of songs is infinite, which also sets it apart from the repertoire of human utterances (or of human songs).

The relevance of these facts to our present enquiry inheres in the fact that species comparatively remote from *homo sapiens sapiens* in their genealogy display the independent development of comparable psychoacoustic abilities. The songs of the passerine birds are made up of combinations of indivisible constituents with constraints on structure, a creative routine entirely analogous to that used in the productions of grammatical speech. Plenty of intra- and inter-species

---

[8] In particular, of course, the definition ignores the other human cognitive 'separation' of auditory input into unstructured noise and the structures of music. Musical perception involves a trip down a cognitive pathway which arguably ends up at a *different* (though not necessarily less 'central') processing level (cf. Jackendoff 1987, ch.12). So though a vital distinction would have to be drawn above the level of 'psychoacoustics' between music and noise perception in humans, this distinction is outside the scope of our discussion.

communication takes place without linguistics, and this can be done with speech sounds if you happen to be a mynah bird. One particular psychoacoustic ability - categorical perception- shared by humans and South American rodents becomes important to our discussion in the following chapter.

1.3.3. 'Phonetic'

At some point we must build a bridge between grammatical objects and structures on the one hand, and the sounds which represent them on the other. It is at this point that opinions begin to be divided. Burton-Roberts & Carr hold that phonetics and phonology are 'inextricably intertwined', and that both remain outside the province of the genuinely linguistic, with the phonology being best characterised as a 'conventional system of physical representation' (at least for speech), 'literally now ... an interface between the linguistic and the phonetic' (Burton-Roberts & Carr 1996: 37). Harris (1996) clearly commits to a position where the fully abstact, linguistic primes of (melody in) phonology each enjoy 'stand-alone phonetic interpretability' (Harris 1996: 551). Life is probably not made easier by definitions like: 'PHONETIC FORM: the output of the phonological component of the grammar, or the phonological component itself' (Crystal 1991: 259). It would be possible to infer from this that 'phonetic' is not commonly regarded as independent of 'phonological', and even that the two may be conflated. '(Phonological) representations have been developed on the basis of the spoken language modality ... (and) are often so close to the phonetics of spoken languages that we cannot rule out the possibility that non-trivial aspects of them are modality-specific' (van der Hulst 1993: 209). This all results in a definition of 'phonetic' that has a mentalistic, or even linguistic, spin to it, and finally confuses the issue by adding an extra, quite unnecessary, perceptual level to a linguistic model: witness 'a consonant lacking a phonological place of articulation is *phonetically interpreted*[9] as a velar' (Rice1996: 494). The best way out of this seems to be to occupy the simple position expressed by the following statement:

---

[9] Emphasis added.

anything phonetic is not linguistic: anything linguistic is not phonetic. This is pretty clear if we look at the fields of study of the branches of the science of 'phonetics'. These are:

(i) *Acoustic phonetics:* the study of the physical properties of speech sound. This is therefore a subdiscipline of that branch of physics which studies sound (i.e. acoustics) when that sound is made by the human vocal tract and *used in the service* of language: there is nothing that acoustic phonetics informs us about linguistically significant *speech* that it could not do for any other sound, and in particular for non-linguistic vocalisations. The objects of study are not linguistic.

(ii) *Articulatory phonetics:* the study of the way the vocal organs articulate *speech* sounds. This is therefore in part straightforwardly a branch of physiology (and anatomy): the specification of places of articulation and all other available phonetic parameters involve physiological (or other physical) mensuration. Once again no specifically linguistic information is delivered by these types of description. However, it may be (wrongly, for reasons to be made clear) argued that articulatory phonetics is also concerned with the transduction of language into the physical plane. We return to this in a moment, because if this were true, the inverse should be true of:

(iii) *Auditory phonetics:* the study of the various transductions that take place between a sound pressure wave and an electrochemical brain operation. Therefore the field of enquiry is once again squarely set within either a physiology, an anatomy or a physics framework. At the far end of this chain of transductions we end up at a point where it is possible to observe such phenomena as timed nerve-firings and ERP[10] changes; still physical

---

[10] Event-Related Potential: see Woods & Elmasian (1986), and for a brief discussion of ERP tests in the present context, see §4.4.2 (p.119).

> events, but ones which may *putatively* be found to be associated with different linguistic activities. There is absolutely no empirical suggestion at present that such an association holds, but even if it did, it would be no more use in characterising the linguistic than noting that English native speakers employ very similar articulations when pronouncing the beginning of the words 'past' and 'pasta', but that these do not correspond in voicing with the pronunciations of an Italian native speaker. In other words, all that is assessable from such information is the quality of the description of a thoroughly extralinguistic event.

In like manner, the fact that sensorimotor processes drive articulatory phonetic reflexes means that we could unpick these transductional processes in the greatest detail without uncovering a trace of anything linguistic. For sure, there is sensorimotor 'knowledge', but the position adopted here is that this 'knowledge' of the position and movement of parts of the organism has not been shown to be language-specific, being as relevant to proprioception as it is to speech.

The three branches of phonetic science, then, do not involve themselves with the linguistic. This does not detract from any contribution a phonetician may make to the statement of systemic language (ir)regularies on the basis of phonetic indications, but such a statement is itself not a statement about a 'phonetic' system, since language-systems are not based on physics or physiology. 'Functional phonetician' is synonymous with 'phonologist'. 'Linguistic phonetics' is an inherently paradoxical phrase. Under this very heading, however, Ohala (1997) summarises 'the relation between phonetics and phonology' thus: 'phonology seeks answers to (mentalistic)[11] questions employing methods from phonetics' (1997: 693). This characterisation of the relationship is entirely consistent with the present approach and does not define or require a type of phonetics that is linguistic.

---

[11] Parenthesis added.

In humans, psychacoustic perception maps to language. What, then, could be meant by 'phonetic perception'? Nothing at all, unless in the special sense of the trained conscious attention that a phonetician pays to the sounds of speech. This attention bears the same relation to native-speaker ability as piano-tuning does to musical enculturation.

One last bastion of 'phonetic perception' may be (and has been) posited in dialectal differences, due to the 'You Like Potato'[12] line of reasoning. In London, the utterance [bɑːθ] refers to the object known in Liverpool as a [bæθ]. Since we know the referents of these utterances to be the same thing, are not [ɑː] and [æ] perceived as 'phonetically' different, but 'phonologically' the same? No, for two reasons. First, there is no qualitative difference between a 'language', a 'dialect' or an 'accent', but simply different degrees of common ground. Knowing that [bɑːθ] and [bæθ] refer to the same thing involves speaking the two relevant 'dialects': in other words, perceiving the phonology of each one. It is easily conceivable that someone who regularly and unconsciously does this may not be able to perform the same trick given an unfamiliar, though acoustically approximately equidistant, pronunciation of the same word (New York City English [bɪəθ], for instance) without the benefit of several additional contextual cues. The second reason why this perception is not 'phonetic' lies in the characterisation of the difference between [ɑː] and [æ]. Presumably, mynah birds from London use the first of these pronunciations and Liverpudlian ones the second. They don't need 'phonetics' to do this, just, in our terms, 'psychoacoustics'. If we answer this objection by importing human 'lexical meaning' into the characterisation of the contrast between [ɑː] and [æ] on the grounds that the two sounds can be lexically distinct in both languages (though not between [b] and [θ]), things just get worse. The result is two indivisible 'phonemes': phonemes have been discredited as formal units for thirty years, and in any case were most definitely not 'phonetic'.

---

[12] [n̩ ˈaɪ laɪk pəˈtʰɑːtəʊ]. From *'Let's Call The Whole Thing Off'* by George & Ira Gershwin (1937).

Our conclusion, then, is that 'phonetic' perception may never be deemed linguistic, but must mean introspection on properties of the sounds of speech using encyclopedic knowledge. This will be important to our discussion of the existing literature on infant acquisition in the next chapter, where it will be obvious that, tacitly or otherwise, researchers think of 'phonetic' perception as some sort of uniquely human precursor to phonological acquisition. Since phonetic perception is conscious attention, and language acquisition is noncontroversially subconscious, phonetic is not linguistic.

From hereon, apart from when it is obvious that the word is set in the context of cited references, in which case it will be marked in quotes, subsequent use of the term *phonetic* in the present work refers only to properties of speech subject to mensuration.

1.3.4 'Phonological'

The *use* of sound in language is not exclusively driven by physiology. Distributional regularities exist that patently come from elsewhere. All human babies produce bilabial trills: no (or next to no) human language uses bilabial trills. All human languages use [i]: some also use [y]: none uses the second but not the first. Tricorn vowel systems overwhelmingly use [a], [i], and [u]. Young language acquirers may harmonise consonants: steady-state languages generally harmonise vowels.[13]All human languages have stops: most have fricatives, too: none has fricatives only. /r/-sounds displaying identical cross-linguistic phonological distributions exhibit a wide range of place and manner specifications.[14] The simplest account of these facts does not suggest that physiology is a primal factor

---

[13] For more on this, see §6.3.3, (p.188).

[14] The diversity of the phonetic manifestations of rhotics is clearly laid out in Ladefoged & Maddieson (1996): after a detailed description of these manifestations, which include approximants, trills, taps and fricatives ranging from dental to uvular in place of articulation, the conclusion is drawn that 'the overall unity of the group seems to rest mostly on the historical connections between ... subgroups, and on the choice of the letter 'r' to represent them all' (Ladefoged & Maddieson (1996: 245).

in the acquisition of phonological systems.

Anything 'phonological', therefore, relates to the mentally constituted system that compels and conditions the sounds (including silences) of a language. This definition should so far satisfy those phonologists who hold that phonological components and relations are purely linguistic in nature. Under this assumption, componential and relational properties can be both in the lexicon and at PF: individual objects may be phonetically interpretable, or may need compounding for this to be possible, depending upon the phonological model that is espoused. As it stands, though, the definition above should also satisfy those, like Burton-Roberts & Carr (1996), who see phonology as a non-linguistic interpretative mechanism between the language faculty and sound, since 'mentally constituted' does not equal 'linguistic'. Some further definitional focus is therefore needed to contextualise the present discussion in the light of these competing points of view.

One respect in which phonology could be said to fail to be 'purely abstract' is in its mapping (however this mapping may be characterised) to sound. The 'pure' concept of phonology could have been more easily maintained if there had been any very successful accounts of the phonology of sign.

It has been spectacularly demonstrated by Kegl and her co-workers that if a community of deaf people arises as a result of historical accident, those among the community who have not yet passed the critical period will spontaneously develop Language *within one generation*. However, all the best evidence is syntactic.[15] Attempts to line up phonology and sign-structure seem more suspect.

During the course of an introductory work on the 'Phonology of ASL', Coulter & Anderson cite as evidential parallels (1) the organisatation of utterances into segments, (2) the existence of a sonority hierarchy, and (3) the existence of co-

---

[15] For a comprehensive list of references to works by Kegl and her colleagues, see publications by the American Sign Language Linguistic Resource Project, Boston University, at http://web.bu.edu/ASLLRP/publications.html. All concern syntactic structure.

occurrence constraints (Coulter & Anderson 1993: 12). 'Segments', however, are not necessarily phonological units, but linearly aligned objects which could as easily be representational of other categorical systems (such as syntax) unless they are further specifed. The best example of a sonority effect that is offered is the behaviour of a sign which looks a bit like a classic 'sonorant' in that it comes after the inital sign of an utterance and before the 'syllabic peak'. However, it then proceeds to be prohibited in just the place where it is predicted to exist in a rhymal closure (op. cit.: 9). Finally, the three examples of co-occurrence constraints that are presented are the signs for 'HEAD', where the signer indicates two points on the head, 'FLOWER', where each nostril is indicated, and 'EYES' where the signer points at each eye in turn (op. cit.: 12). These signs are just too full of iconic content to be considered arbitrary, in the way that the sound-address of a lexical item is arbitrary. They are therefore not easily comparable to phonological objects. Iconicity has certainly a role to play in the genesis of lexical items,[16] but given that there is no such thing as a primitive language, there can be no prediction that the 'phonology' of sign languages should be more iconic than that of spoken language.

Again, in a dissertation which aims at 'laying the groundwork for a model of visual phonology', Uyechi (1996: 17) finds that 'at a low level of representation the constructs of a spoken language phonology are different from the constructs of a visual phonology'. The hope is held out that 'the organisation of these atomic units into more complex units' ...may... 'follow general principles of organisation that are the laws of a theory of universal phonology' (op. cit.: 17). If this is phonology, then it lacks primes altogether, since there is no common ground between the units of 'spoken' and 'visual' phonology. 'Phonology' is left with only 'general principles', and thus becomes a system not easily recognisable to phonologists of spoken languages.

Other attempts to conflate phonological constituents with those of sign language

---

[16] Katamba (1998), for instance, identifies a conflict between sound symbolism and a constraint against voiceless labials in Luganda and uses this to illuminate the origin of certain alternations (in an Optimality Theoretic framework).

seem to end up emphasising differences, rather than similarities. Considering the centrality of the onset constituent to *all* versions of syllable structure theory (even theory without constituents), it would be expected to find some analogous constituent in sign. It is somewhat surprising, then, to find that in a detailed discussion of the units of sign, the claim is made that 'syllables in sign structure lack ONSETS at the phonological level' (van der Hulst 1993: 211). This analysis is interpreted (in the same volume) as 'there are no phonological onsets, ... but only phonetic onsets, which are the transitional movements that precede signs' (Corina & Sandler 1993: 200). Using the definition of 'phonetic' proposed in §1.3.3, it becomes impossible to assess such a statement.

Van der Hulst also notes that 'monomorphemic signs are typically monosyllabic' (op. cit.: 235). Again, such isomorphism is far from commonplace in the spoken-language phonology which will be under consideration here.

Given that visual and aural perception depend on quite different (spatial and temporal) cues, it is an eminently reasonable exercise to see if it is possible to characterise the organisation of sign as 'phonological'. But, from the examples cited so far, the attempt seems to have failed. Stronger evidence of a phonology without psychoacoustics may at some stage be presented. For the moment, however, some ground must be conceded to the 'non-linguistic' school of thought on the nature of the 'phonological', if 'linguistic' means 'irreducible to language-external principles'. Phonology, unlike syntax, must map to the perception (and production) of sound, so users of sign language are never, as far as has hitherto been established, going to need phonology. Let us propose, then, that phonology is 'substrate-specific', founded in sound both phylogenetically and ontogenetically. Can it still be purely abstract? Dennett (1995:149-186) gives an up-to-date summary (for non-specialists) of the state of current knowledge about the origin of life on the planet, and in particular he draws on references which show how simple risk-taking algorithms may have led to the appearance of pre-life forms (like viruses) in a crystalline environment, and eventually to (comparatively) highly organised biological life-forms, 'Vastly' well 'Designed' (in Dennett's terms) to resist entropy. Importantly, for life or for semi-life (organisms which may not live independently), biological algorithms are all that there is. During their mechanical progress through evolutionary pathways, these algorithms may produce systems which exhibit design features possessing abstract properties. 'Flowers are biological systems, but their petals are ... often organized into patterns elegantly generateable from a nonredundant base' (Brody 1997: 8). Obedience to nonreducible abstract constraints is not inconsistent with biological origins. If *syntax* is linguistic, in that it is 'perfect', or 'nonredundant' or 'abstract' in its conformity to principles which are not language-external, and since Language did not evolve by Magick, then a linguistic system can and must be instantiated by a bunch of dumb biological algorithms. An infinitely closer link in the evolutionary chain can be invoked by proposing that the mentally-constituted system of phonology has evolved using the perception of sound, and is acquired by each human by mapping sound perception onto this evolved system. Understood like this, anything 'phonological' relates to a genuinely abstract, and dedicated, system of pattern recognition which depends on the speech signal for its physical and psychophysical (psychoacoustic) correlates.[17] Maybe the fact that the evenly-spaced formants associated with a neutral tube configuration regularly cue *phonologically* neutral vowels is a vestigial ghost of the evolutionary process, rather than a mere coincidence.

Independent systems must be independently acquired. A little evidence for the lack of isomorphism between morphosyntactic and phonological domains has already been presented (§1.1, p.16). This does not prohibit language phenomena which depend on coterminous domains, merely makes them optional. Neeleman & Weerman (1997: 129ff.) use Selkirk's (1986) proposal that phonological phrases are closed at the right edge of XP, where XP is any syntactic phrase, in an account of the setting of a word order parameter during acquisition. They contend that VO languages (e.g. English) differ from OV languages (e.g. Dutch) in that the case

---

[17] More on this in §6.2, (pp.176 ff.).

checking domain of the first type is Selkirk's phonological (ϕ) phrase, and of the second type a syntactic phrase (m-command domain). This is relevant here as circumstantial evidence for the (relatively) close connection between phonology and sound for the following reason. Among the raft of effects produced by the setting of this parameter, the ϕ phrase-setting results in a more constrained linear order in the VO languages than that exhibited by the OV languages: intervening adverbials are permitted only in the latter. The phonological option for the parameter results in a comparatively greater surface adjacency effect being observable.

This type of adjacency effect does not undermine the entirely mentally constituted nature of 'phonological' as it is proposed here. Nor, of course, do our proposals say anything at all about the 'syntactic' *per se.* But the word that *is* becoming etiolated in definitional power is 'linguistic', unless it be used as a generic term covering the various computational systems utilised in the communication of Language. There may be parallels to be drawn between one or another phonological or syntactic model ('Government' is, or was, one), which reflect the operation of language-dedicated principles. But PF must continue to interface directly with the physical, so phonological primes maintain an ancient taproot into sound. This effectively removes any way of assessing the 'phonological' as '*wholly* linguistic', in the way that syntax may be assessed, using 'linguistic' in its strong sense of 'non-reducible to language-external principles'. Finally, the whole exercise becomes meaningless if 'syntactic' is conflated with 'linguistic' in its weaker, generic sense.

Whither modularity, if all this is so? Do we have to live with the fact that we have potentially let the strategies of 'general cognition' into the language faculty by the back door? The understanding of a 'mental module' in this discussion is that it is delimited and defined not by the (generalisable) strategies available to it, but by the designed and specialised summation of those strategies in the service of a particular cognitive process. Anything 'phonological' refers itself to an abstract pattern recognition system that interfaces with psychoacoustics and with morphosyntax.

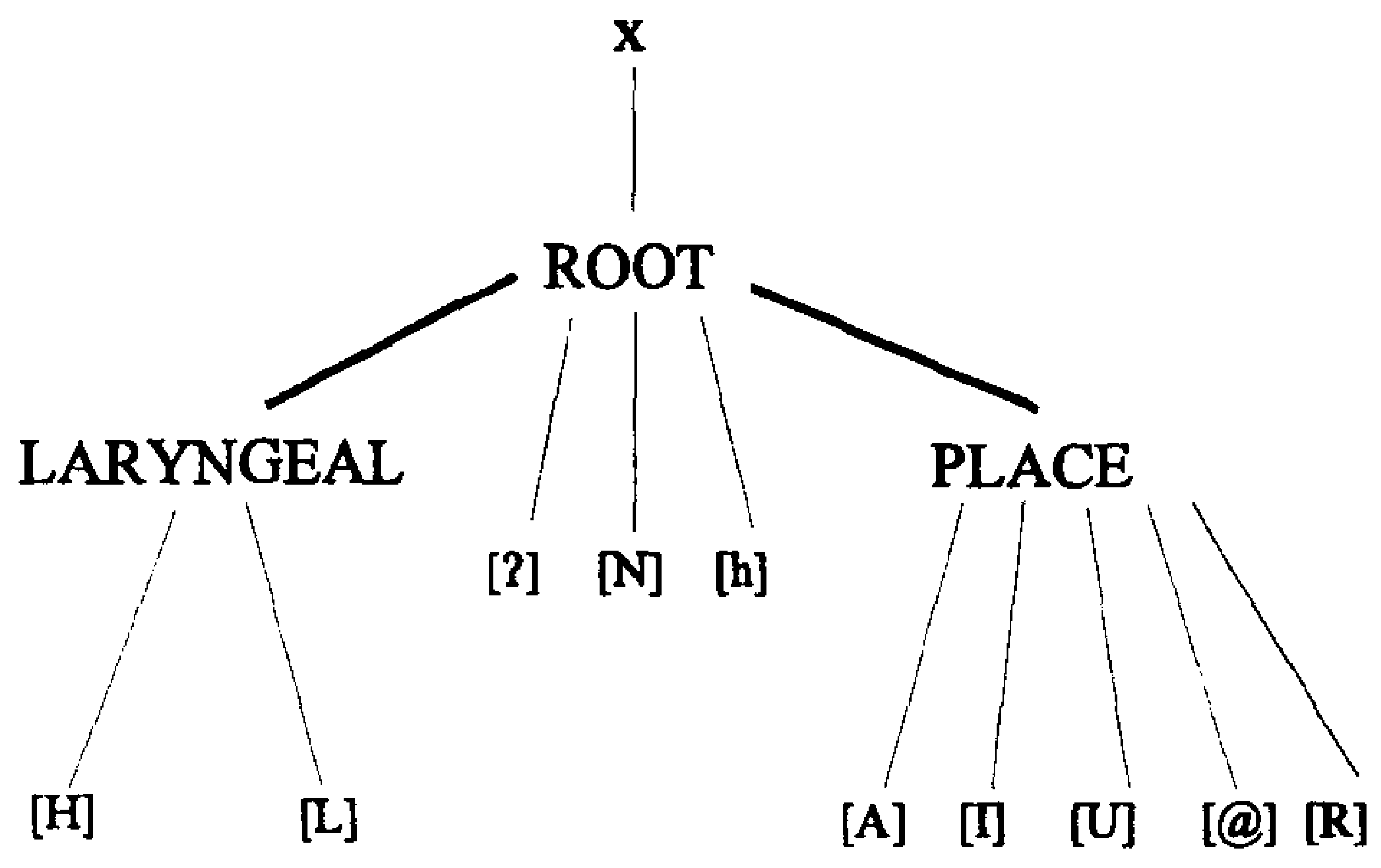
The last part of our definition of 'phonological' needs to repeat that phonology is not *itself* a single system. The independence of prosody and melody was alluded to in §1.1 (p.16) and is supported with examples in §2.3.2 (p.50). Hence phonology should be regarded as bi-modular, rather than as a single module.

1.4 Elements and geometry

The melodic component within this bi-modular system will be referred to frequently in the subsequent discussion and some general aspects of its acquisition form the focus of chapter 6. It is therefore worth setting out the assumptions made here about its internal structure. The system adopted is inherited from Anderson & Jones (1974), Kaye, Lowenstamm & Vergnaud (1985) and Schane (1984) among others, and specifically as it is expounded in Harris (1994) and Harris & Lindsey (1995). Melodic material is analysable into a small set of primal 'elements', which have a psychological reality as patterns: these patterns are reflected in aspects of the acoustic signal. Elements are entirely privative, are autosegmentally arranged on tiers, and may be independently phonetically interpreted, or may synthesise to form complex melodic expressions. Head/dependent (also referred to as head/operator) relations within these complex expressions are used to account for some distinctions previously regarded as 'phonemic'. The linking of elements to prosodic structure underlies processes such as spreading and assimilation, and their delinking to account for lenition trajectories and diphthongisation. Elements are affected as classes by phonological processes, and infrasegmental geometry therefore fulfils the same function in this element-theoretic model as it does within a feature-theoretic approach.

A diagrammatic representation of the nodal association of the ten elements which, in Harris (1994), are the maximum (and minimum) considered necessary to generate melodic expressions is presented in (3).[18]

(3)



A further element often referred to in the literature, [ATR], is missing from this inventory. This element is a focal point for the discussion in Kaye, Lowenstamm and Vergnaud (1985), where it is used to account for the contrast between 'tense' and 'lax' vowels. Other accounts have done without this particular object. Schane (1984), for example, proposes that such contrasts are underpinned by instances of duple or single elements. Harris (1994) adds the responsibility for the 'tense/lax' contrast to the workload of element [@]. Element [@], which is equivalent to Kaye, Lowenstamm & Vergnaud's 'cold vowel' (v), has no tier, but is regarded as a background whose signature will only become evident if it heads an expression, or if it is the sole participant. A lax vowel will be headed by [@]: the representation of English /u/ is thus [U] and that of /ʊ/ is [U@].[19] The existence of [ATR] as an element is additionally undermined by its distinctly un-elemental property of being

---

[18] For the acoustic correlates of these elements, see Appendix A and Choo & Huckvale (1993: 341).

[19] As is customary, elemental complexes are written head-rightmost in running text.

phonetically unrealisable in isolation.[20]

Various adaptations of this inventory have been proposed, most of which have either the aim or the consequence of reducing it.[21] Besides the necessity of partitioning elements into natural classes, it has become obvious from these reductionist endeavours that some elements are more equal than others. 'Manner' elements [?] and [h], for instance, unlike place and laryngeal elements, do not participate in long-distance relationships. A model which includes melodic headship (such as Ritter 1997) may be able to dispense with these elements altogether. In §5.2.2 (pp.134 ff.), the question of the viability of conflating [L] and [N] is also raised as it becomes relevant to a discussion of Yorùbá tonology. It is worth jumping the gun here to note that if this conflation were to be accepted, all widely-attested local and long distance phonological distributions and processes (save stricture which may putatively be linked to headship) become the responsibility of five elements. These naturally divide into two groups, and their territories of operation are laid out in (4).

(4)

Elements	Single-position contrasts	Long-distance correspondences
A I U	Vowel colour in nuclei Place contrasts in onsets	Vowel harmony
H L	Phonation contrasts in onsets Nasality in onsets and nuclei Lexical tone in nuclei	Intonation Nasal harmony Tone copying, spreading

---

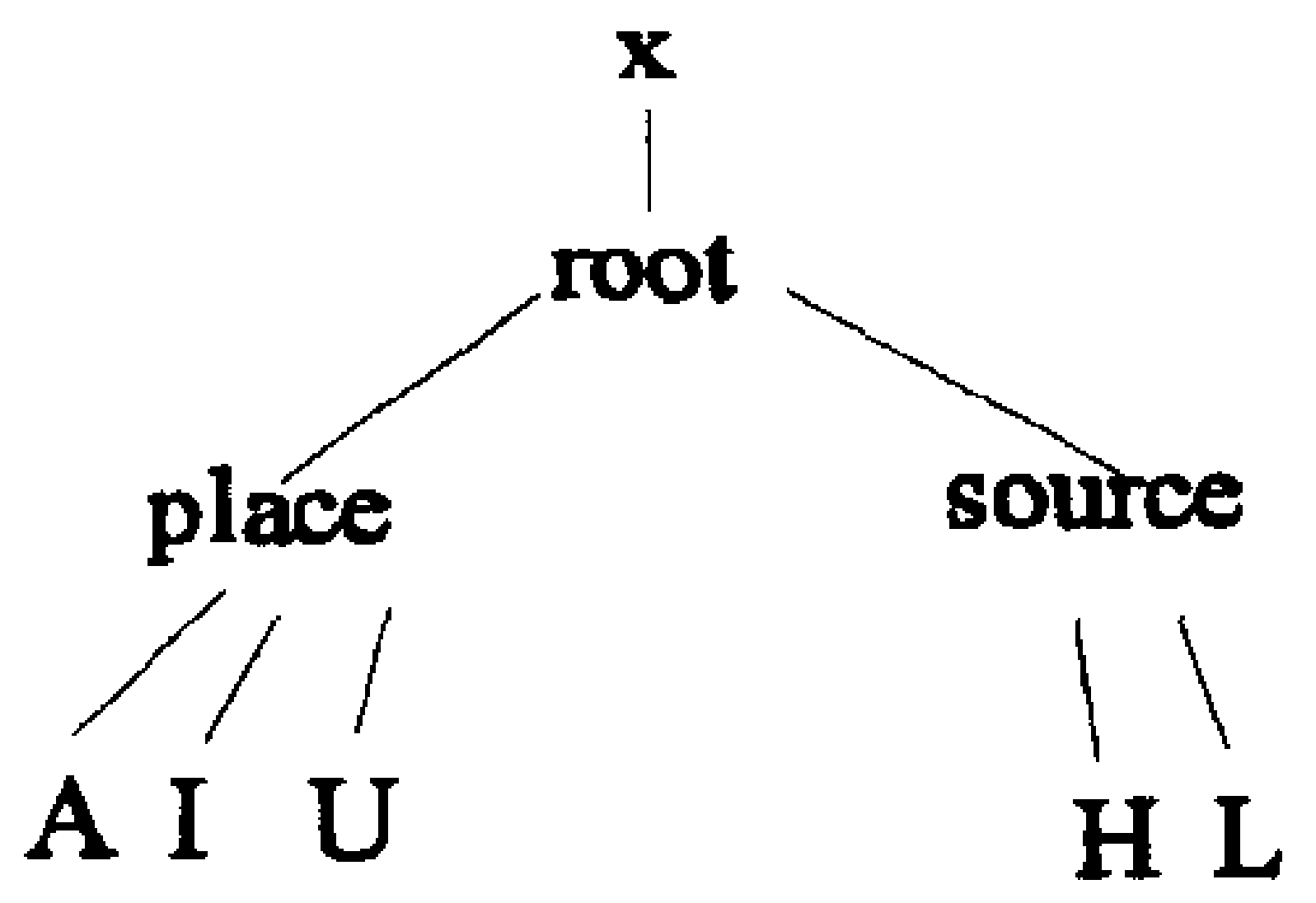
[20] Kaye *et al.* (1985) claimed that it was pronounceable, and they did so on the theoretical grounds that it, 'like the other elements of vowel systems...is a fully specified feature matrix' (1985: 312). Using data from Kpokolo, they went on to argue empirically that its realisation is a 'tense' version of [@]. Given that there are no features in the model we adopt here, their theoretical argument cannot be fitted in, but a 'purely' elemental theory may be able to account for Kpokolo in terms of headship, just as it has for English.

[21] For example, van der Hulst (1994), Cabrera-Abreu (1996), Ritter (1997), Nasukawa (1998). All these sources will be exploited in the following chapters for various reasons.

For some of these processes, there are obvious pairings between single-position and long-distance operations. In conformity with previously cited sources (Harris 1994 etc.), we accept that this is best represented by stating that long-distance melodic relationships are contracted via prosodic projections which are adjacent at higher levels.

With the reduced elemental set, can we now subtract one node from the three proposed in (3) above? The proposed removal of [N], [h] and [?] will reduce the motivation for a 'Root' node, but not remove it entirely. The vestigial reason for its retention is its function as 'the matrix which defines the integrity of the melodic expression' (Harris (1994: 129). Operations which insert or elide entire melodic expressions need to be represented by the linking or delinking of a matrix node. So reductionist geometry ought to maintain the configuration in (5).

(5)



Both (4) and (5) make it blindingly obvious that [A], [I] and [U] form a natural trinity and [H] and [L] a duality. There is an overwhelmingly robust collection of phonological data which attests to this sub-grouping. It is supported from a psychological point of view in §6.2 (pp. 176 ff.). This fact should be made plain in the theory, and by persisting with infrasegmental geometry involving 'place' and 'source' nodes in any reductionist model, this can be achieved.

These issues will often be raised again in the succeeding chapters.

1.5 A note on serendipity and coherence

The work described in the following chapters took place over a three-year period. In the interests of a coherent presentation, it is not presented chronologically, but rather each chapter represents a 'topic' on which the discussion is focused. Chapter 2 is the only one which genuinely predates the others as it reviews the literature which inspired the investigation in the first place. After that, all the chapters have been written and re-written in the light of each other. Not that this is an unusual approach, but it does bear mentioning in this case, as some of the answers presented relate to questions whose existence was not foreseen at the outset. In particular, the idea that nuclear element [L] and low-tone may not be mapped onto each other in Yorùbá (and other west African languages) was not a focus of the initial program, but arose from the serendipitous and unexpected results of the adult perception tests described in §4.2.6 (p.106). These tests were originally aimed at seeking a chronological parallel in acquisition between two phonetically disparate objects, but even before getting as far as infant perception, the anomalous adult findings concerning low-tone demanded attention, and suggested urgent recourse to work on the tonology of Yorùbá. This recourse turned into chapter 5, and did cut back the time available to fully explore other envisaged avenues, particularly the opportunities for infant research suggested by the results presented in chapter 3. It is proposed here that any incompleteness which has arisen as a result of succumbing to the lure of serendipity has been more than compensated for by the potential links forged between perception, typology and phonology in the discussion of Yorùbá.

In any case, the only 'completeness' that is aimed at or attainable by the present work is to completely convince the reader of the phonological validity of infant perception. This ought to render the conspectus of extant acquisitional perception tests 'incomplete' in the sense that it will be possible to go on to inform both existing and future research with the notions of phonological theory.

PAGE
MISSING
IN
ORIGINAL

Chapter 2

Phonology and Infants' Perceptual Abilities

2.1 Introduction

A good deal of painstaking empirical research is detailed in the literature on language acquisition in the first year of life. It focuses, for obvious reasons, on the infant's perceptual abilities rather than on production, and reveals what is perhaps an initially surprising sophistication in a young child's expertise at discerning contrasts in the acoustic signal which coincide exactly with those used in adult language. In this discussion we will survey some germane findings relating to children of around six to eight months, with the principal aim of relating these findings to general phonological theory, and show that the examination of this relationship inspires research in infant perception. It is an important factor in any account of acquisition that it takes its place within a general theory, though this relationship seems not to be a primary focus of much of the literature we shall be drawing on: it is perhaps worth mentioning, as Menn (1980) has done previously, that Roman Jakobson, whose work advanced many notions that have been drawn upon in modern phonological theory, regarded his highly influential ideas on acquisition as essentially a demonstration of the power of his general theory.

Quite often in the acquisitional literature, the ascription of precocity is where the story ends. This applies both to the acquisition of language and to other developmental accomplishments (facial recognition being a case in point). 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 whether such abilities are present, but how they are possible and what cognitive processes are involved' (1995: 1307).

We review in sections 2.2.1 and 2.2.2 some of the findings of Eilers, Eimas, Kuhl, and their co-workers during the past twenty-five years relating to the perception of segmental contrasts in early infancy. The infant perceptual abilities illuminated by these researchers have, we note, been characterised as 'merely phonetic'. This view has led to some difficulties in aligning their findings with those of other researchers within a single acquisitional account. In §2.2.3 we go on to review these potentially paradoxical results, as reported by Mehler, Jusczyk and others. In §2.3 we propose a generalised theory of phonological acquisition which offers a solution to this contradiction. Section 2.4 gives a specific research proposal inspired by the discussion, and we summarise in §2.5.

2.2 Empirical findings relating to the perception of speech in early life

2.2.0 The phonetic literature abounds with tales of infant precocity. Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison (1988) have shown that neonates can distinguish their native language from timing and intonation cues alone. Russian, Italian, English and French speech signals, from which all acoustic information above 400Hz had been removed, were played to four-day old French and American babies, and their interest in the input measured using the High Amplitude Sucking (HAS) method[1]. The filtering effectively removes all segmental information while leaving the prosodic shape of the stimuli intact. The results confirm that infants prefer their own prosody to that of other languages, while failing to distinguish between foreign prosodies. It is quite plausible that humans begin to attune their native prosodic perception prenatally, as being cocooned in intrauterine fluid would have a similar low-pass filtering effect to that employed in the experiment. It is arguable that the perception of segmental contrasts, on the other hand, has to wait at least until we emerge into life in the atmosphere. But several researchers have suggested that it is not that long after birth until we begin to do so.

---

[1] A pacifier (dummy) is attached to a device that measures sucking rate. Greater attention is accompanied by a faster rate, habituation by a slower: see Jusczyk (1985).

2.2.1 The categorical perception of stop contrasts by babies, chinchilla and students

We now recapitulate the results of some experiments in the categorical perception[2] of stop consonants. Eimas, Siqueland, Jusczyk & Vigorito (1971), again using HAS, detected a Voice Onset Time[3] (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. However, any hopes that may have existed that this could be characterised as 'innate linguistic perception' were quashed by series of experiments reported in Kuhl & Miller (1975) which demonstrate that the South American chinchilla can pull off exactly the same perceptual feat, and at around the same point in the VOT continuum. Categorical perception, though utilised linguistically by humans, is evidently part of a shared mammalian endowment.

Not all languages that utilise 'voicing' distinctions cued by VOT values[4] 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[5], 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.

The coincidence of this finding with adult perception and the extension of this pattern to other language groups is confirmed year after year in a brief experiment carried out at

---

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

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

[4] In the mature language, VOT is not, of course, 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.

[5] Described in detail in §4.3.1 (p.101).

University College London using postgraduate students as subjects. The students come from diverse linguistic backgrounds and are asked to produce intervocalic tokens of three 'voiced' and three 'voiceless' stops, at different places of articulation. The speech wave of each production is compared with a laryngograph trace, and VOT is therefore readily discernable. A typical result[6], illustrated by the graphs in Figure 2.1, shows genealogy-independent setting of the VOT categorical boundary, with Cantonese, Persian and English using a lag value of 30+ms, while Japanese patterns with Spanish, Italian and French in its employment of the 0ms value.

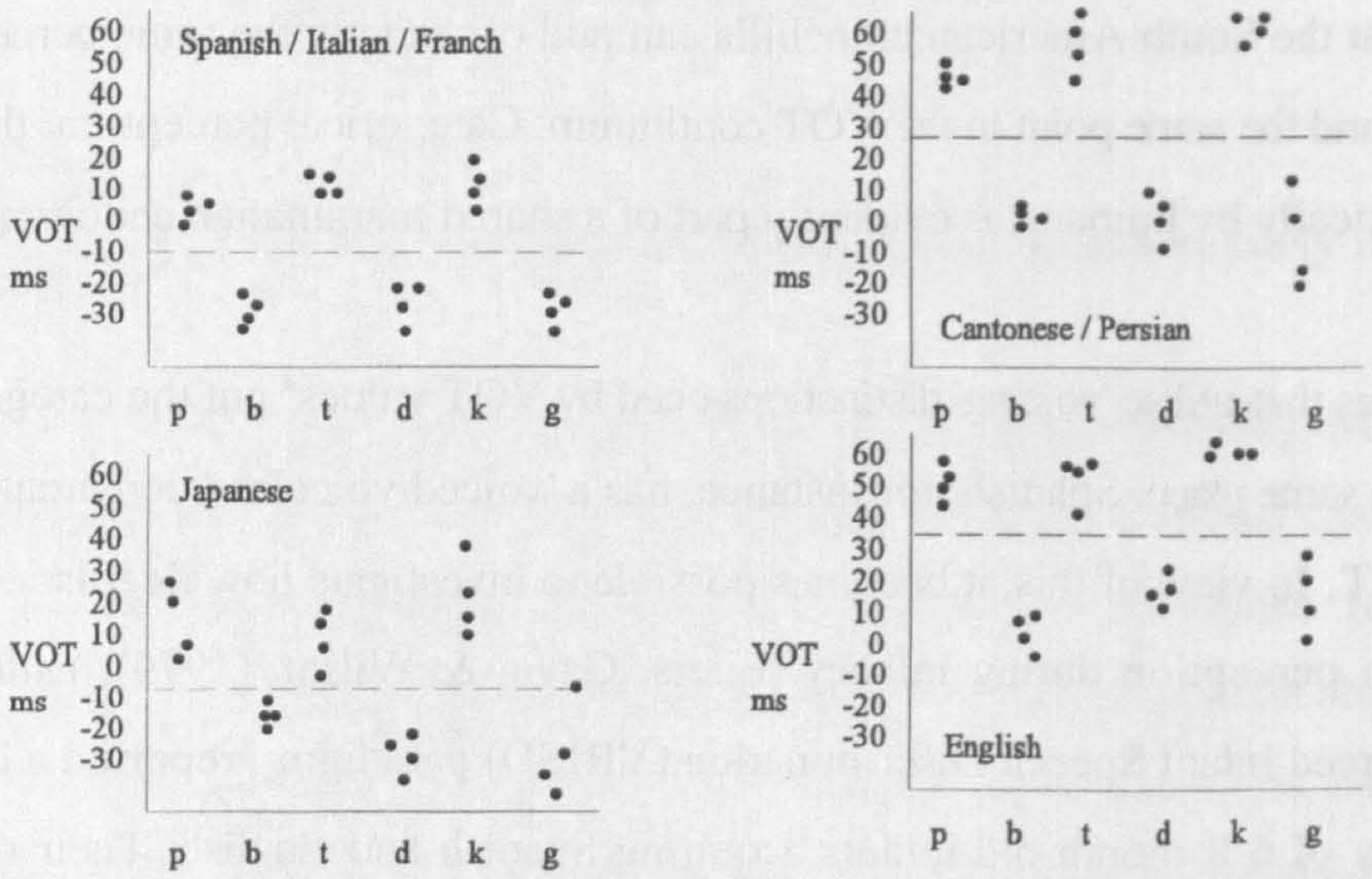


Figure 2.1: Categorical production of stop consonants by four groups of students from diverse language communities (N = 16).

It is given to *homo sapiens sapiens* and to others to discriminate categorically: the factory settings for both human and chinchilla compel the organism to do so at a value of approximately +25ms lag from release of closure. However, human ontogeny and reinforcement from linguistic input may cause perception of this type to be switched to a value elsewhere in the VOT continuum. It is worth noticing that the VOT value of the endowed boundary coincides with the most highly-coloured acoustic region in the continuum. A look at spectrographic analyses of these types of stimuli reveals that the

---

[6] This particular test took place in February 1997.

Spanish boundary has for its acoustic correlate only the presence or absence of a low-frequency voice-bar, but the English boundary is bang in the middle of a region where the first-formant transition is being progressively attenuated and the excitation of the upper formants is also altering. However, the lack of acoustic salience at 0ms VOT does not mark this value out as typologically marginal. Its use for linguistic contrast is widely attested in genealogically diverse languages including Japanese, Dutch, Greek and Hungarian. Considering the acoustic bias to use the given demarcation value, this is evidence that the acquisition of a linguistic sound system is 'unnatural' in the sense that Anderson (1981) uses the term with regard to the steady-state and to diachronic change. Mechanical (auditory or articulatory) conditions do not determine aquisitional results, even if they may be one constraint on development. Though we may trace certain acoustic 'jumping-off points' for some developmental processes, the language-specific acquisition of stop contrasts by Spanish and English infants constitutes evidence that mechanics are only one factor in the matter.

2.2.2 The prototypical perception of vowels by infants

A different perceptual strategy comes into play in the consideration of vowel contrasts. The principal acoustic cues for a vowel are its formant values, and by far the most important of these are the values of the first and second formants: recognisable synthetic vowels can be manufactured by manipulation of F1 and F2 only. Tests on adults show that a native-language vowel is established at a prototypical point in psychophysical space, which may be measured two-dimensionally as an intersection of F1 and F2 values.[7] Tokens close to that value are adjudged 'better' instances of the vowel than those further away. A prototype affects acoustic perception by warping perceptual space so that vowel tokens close by the prototype are more difficult to tell apart than those at some remove (Grieser & Kuhl 1989, Kuhl 1991).

The important point for the present discussion is that Kuhl's acquisitional studies (Kuhl 1991, Kuhl, Williams, Lacerda, Stevens & Lindblom 1992) indicate that this warping has

---

[7] The role of normalisation is ignored here: we discuss it later in this chapter and in chapter 3.

already taken place by the age of six months, the same age by which native-language input may have adjusted VOT perception. There is a parallel compulsion to ignore acoustic equidistance in favour of linguistic ecological value. The Swedish/English cross-linguistic study recounted in Kuhl (1992) begins with a verification test on native speakers showing Swedish /y/ to have a typical F1/F2 plot of 300/1550 mels, while that of American English /i/ is 350/1900 mels. Thirty-two variants surrounding these prototypes were synthesised, and, using VRISD on infants of 6 months, the results showed that Swedish babies perceived the /y/ variants as identical more often than chance (66.2%) with a result much closer to random in their perception of the foreign prototype (55.9%). The findings from the American children were more impressive still: 66.9% for /i/ against 50.6% for /y/. Just as in adult mentation there was a loss of perceptual finesse close to the native prototype.

Kuhl has modelled this developmental story under the banner 'Native Language Magnet' (NLM) theory (Kuhl 1992, 1994).

Figure 2.2: Diagrams from Kuhl (1992)

Our mammalian inheritance is represented in Figure 2.2(a). This map is quintessentially conceptual and could be drawn in any way that 'conveys the idea that at birth humans can discriminate all phonetic (vocalic)[8] contrasts' (Kuhl, pc). So the plot would presumably be just as valid with F1 on the ordinate and F2 minus F1 on the abscissa, and we should be careful not to posit an overly simplistic relationship between these representations and the parametrisation of the primes which can be active in structural relationships. The nature of the 'psychoacoustic vowel space' is explored in Lindblom (1986). Here, calculations are presented which take account of basilar membrane mechanics, non-linear frequency response and finally utilise a loudness calculation to deliver a true auditory map. We take this warning against 'formantism' seriously (it is repeated by Disner (1986) in the same volume) but as long as Kuhl issues the 'concept only' rider alongside her diagrams, the two views do not seem incompatible.

NLM theory proposes that humans develop a finite number of perceptual vowel targets, each of which will become located within one of the subdivisions in Figure 2.2(a). These are prototypical in nature, and each succeeding exemplar around such a 'magnet' will reinforce its psychoacoustic reality. Language input, therefore, turns us on to the magnets that are present in our native language using a neurological response to input stimuli well ahead of any requirement for word recognition. By the age of six months, Swedish, English and Japanese infants have become native-perceivers of their own language to the extent that they are responsive to only one of the three vowel prototype inventories in Figure 2.2(b). If we posit a one-to-one relationship between NLMs and mature vocalic contrasts (a point to which we return in §2.3) then, as presented here, NLM theory predicts a universal maximum number of vowel quality contrasts in human language, derived from the endowed subdivision of psychoacoustic space. The thirteen Swedish vowels (and the thirteen vowels of the Amstetten dialect of Bavarian reported in Ladefoged & Maddieson (1996: 290) exploit all the possibilities, but the attestation of a larger inventory, such as that proposed for Norwegian in Maddieson (1984), does not weaken the concept; it just changes the map.

---

[8] Parenthesis added.

What of the responses to vowel distinctions which are possible developments but are irrelevant for the infant's native language? NLM theory predicts that without reinforcement, these will atrophy. With the development of targets at particular values, some of the boundaries that divide the vowel space become irrelevant, and the ability to perceive them at all eventually disappears, as represented in Figure 2.2(c). It is biologically commonplace that the developing organism responds to environmental stimuli at critical windows of opportunity, and if such an opportunity is missed, it is missed for ever. Thus in independent research, it has been demonstrated that there is a decline in sensitivity to foreign-language contrasts in the first year of life. Werker & Lalonde (1988) present results that show that at six months, English-acquiring children can distinguish [ɖa] from [da] and [ta], but that this ability is gone by one year: it continues to be present in Hindi-acquiring children for whom the contrast has linguistic significance. Best (1991) shows the same decline in sensitivity to the Zulu lateral fricative voicing contrast (/ɮ/-/ɬ/) in English babies.

The presence of NLMs is also said to actively interfere with the perception of foreign-language contrasts in later life. A non-native contrast will be more difficult to perceive if the distinction lies acoustically close to an NLM. Thus the acoustic proximity of the Japanese native /r/ or Korean /l/ to both English /r/ and English /l/ explains why this should be a classically 'difficult' foreign distinction to perceive. Tsushima, Takizawa, Sasaki, Shiraki, Nishi, Konho, Menyuk & Best (1994) show that by the age of one year, the [r]/[l] contrast is already being misperceived by Japanese children, though this particular contrast is still perceivable at six months.[9]

There are further strategies used in the perception of speech sounds which are identical in infancy and adulthood. Kuhl's research has elsewhere demonstrated that infants employ normalisation (Kuhl 1979, 1980); they are able to determine the identities of spectrally dissimilar vowel colours independently of pitch contour or speaker. Experimentation in

---

[9] A contextual setting which is independent of the various research endeavours cited here is that sensitivity to vowels seems broadly to predate sensitivity to consonants (Werker & Polka 1993), a finding that Cutler & Mehler (1993: 106) attribute to perceptual factors which may be collectively termed 'salience'.

cooperation with Meltzoff (Kuhl & Meltzoff 1982) has also suggested that infants, just like adults, will use other perceptual modes in speech processing; the 'McGurk' effect, which demonstrates that visual information can interfere with speech recognition in adults (Summerfield 1983), is anteceded by the conviction in infancy that certain vowel sounds go with certain facial configurations. Since both bimodality and normalisation are entirely usual strategies in general perception, it is logical that they should take place independently of the specific computations which relate NLMs to a developing phonology.

How this latter relation is accomplished is, as we have said, not yet clear, but it is a valid question to ask, and it can be asked validly only by being lucid about the entities we are trying to relate to each other. We take this matter further in §2.3.

2.2.3 NLM theory, WRAPSA, and the size of infants' units of perception

Bearing in mind the perceptual finesse evinced at six months of age and proposed by Kuhl's theory, we now question how this may tie in with the size of the segments perceived. There is a good deal of research which helps to establish the idea that a major early 'gateway' to parsing is the prosodic structure of the native language, and that the first discernable units to be perceived are suprasegmental.

In addition to the work already cited in §2.2 (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison 1988), more evidence of the relevance of prosodic information to early acquisitional strategy is presented in Jusczyk & Derrah (1987). They demonstrate that, after habituation to a series [bi], [ba], [bo], a two-month-old infant will dishabituate equally to *both* [bu] and [du]. The same result has been shown in Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler (1988) for vowels: habituation to [bi], [si], [li], [mi] is followed by dishabituation to both [di] and [da]. These findings have been interpreted as indicating that while the 'syllable' is discernable, infants do not recognise 'phonetic' similarity. Other laboratory work by Jusczyk and colleagues bears results that demonstrate sensitivity by nine-month-olds acquiring English to the dominant trochaic stress pattern of that language, a sensitivity *not* found by these investigators in six-month-olds (Jusczyk, Cutler & Redanz 1993). Thus Jusczyk, within the framework of his Word Recognition and Phonetic Structure Acquisition (WRAPSA) model, proposes the development of a 'weighting' scheme from the age of six months, which filters the information available from general auditory analyzers; pattern extraction from the output of this system leads to the representation of 'syllable-sized' units, and these representations form the gateway to phonological organisation.

There seems, therefore, to be some discrepancy between the results achieved by Jusczyk and those achieved by Kuhl, the divergence of their views being clear from their two articles in Volume 21 of the *Journal of Phonetics* (Jusczyk (1993) and Kuhl (1993)). Here, Kuhl argues that the part of the discrepancy that is not due to developmental differences because of the subjects' different ages is due to methodological differences. The high-amplitude-sucking (HAS) technique used by Jusczyk focuses on infants' ability to perceive novelty, and the tests he cites show only that differences are perceived: it does not follow that similarities are not perceived. Equally, as Eilers, Wilson & Moore (1977) point out, negative results using HAS are ambiguous, as a failure to dishabituate may simply indicate a lack of interest on the part of the subject, rather than a lack of discrimination. By contrast, both Eilers and Kuhl have advocated and used the visually-reinforced infant speech discrimination paradigm (VRISD). This technique, to be described in detail in chapter 4, involves rewarding the infant for a 'correct' response, so subjects can be reinforced for detecting similarities. Additionally, since the reinforcement is now independent of the auditory signal and under operant control by the infant, a failure to respond is unlikely to indicate anything other than a failure to discriminate.

It is well argued that a visual reinforcement technique is an improvement on HAS, or indeed any other technique in which the auditory signal is both stimulus and reinforcer (such as monitoring heart rate), and the dichotomy between the independent findings on two- and six-month olds may be accounted for by this, taken together with developmental change in the intervening period. But neither Eilers, Kuhl nor anyone else dispute the presence of some validity in Jusczyk's results. In particular, Kuhl (1993) concurs with the importance of suprasegmental representations, merely proposing that it is too early to rule out the segment, 'as Jusczyk seems to recommend: ...a magnet effect at the level of the segment suggest[s] that by six months of age segments do constitute *a* unit of analysis' (Kuhl 1993: 135). An equal willingness to leave this matter open can also be found in the opinions of those whose work has led them to support the suprasegmental position: so Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison (1988) conclude that, while their results using a low-pass filter show that prosodic clues alone are sufficient for discrimination, they do not show that melodic clues alone are not.

Despite these authors' willingness to remain open-minded about such considerations, there does seem to be consensus among researchers working in this field that the attested perception of 'phonetic'[10] segments in the first six months of life does not constitute a phonological system. Kuhl, for instance, despite her results, argues that the development of NLMs is ahead of, and separate from, the development of phonological contrast. She proposes that though the subsequent development of such a contrastive repertoire may cause NLMs to 'repel each other to achieve further distance between them' (Kuhl 1993: 133), the existence of NLMs at six months old is a result of 'attraction or assimilation' only (op. cit.:132). This is assimilation in the Piagetian definitional sense of 'filtering or modification of the input'.

2.3 The relationship between early acquisition and phonology

2.3.0 It is self-evident from the previous section that the acquisitional attainments of the first six months are crucial to the child's development of phonology, and furthermore are language specific, arguing for a direct and simple connection between these attainments and steady-state language. The existence of a continuity between early perceptual objects and mature phonology also squares well with the struggles experienced in learning a foreign language. No matter how much alien vocabulary or syntax we have, the last bastion of influence that a native language yields is the pronunciation and perception of

---

[10] In §1.3.3, 'phonetic perception' was argued to be a potentially confusing misnomer, but since the literature focused on here uses this term, it is repeated without further comment.

sound. The persistence of transfer errors in this area argues for the ingrained nature of an early-acquired sound system.

2.3.1 Infancy and adulthood: continua or transductions?

For what reasons do researchers maintain the *lack* of attachment between early 'phonetic' perception and phonological development? One problem in assessing the status of the representations at six months of age seems to be the accordance of psychological reality to the results of Jusczyk and others with older infants, as discussed in the last section: if some suprasegmental categories are only developed after six months, and if these are the gateways to the perception of phonological representations, then it follows that infants' earlier categorical perception must be qualitatively different from phonology. 'Linguistic contrast' could only be achieved when perception had narrowed its focus from larger sections of input to something 'segment' sized. But since the inventories proposed by Kuhl at six months old are manifestly crucial to the phonological categories of the eventual adult language, there seems to be an obligation on the part of any theory of development to include some account of a continuum between these two, or at the least an interface where one is predictably transduced into the other.

2.3.2 Unchained melody

An essential part of such an account could be achieved by the acknowledgement of the independent development of prosodic and melodic representations, an idea entirely concordant with the phonological theoretical basis we are adopting. This bifurcation within the phonological component can be succinctly illustrated from three independent areas of research which we now recapitulate.

Firstly, speech errors in on-line processing generally conform to predictable patterns, a common example of which is to be found in the incidence of 'spoonerisms'. The mispronunciation of *crushing blow* as *blushing crow*, involving the swapping of the melodic content of the onsets of two phonological words, represents a more attested pattern than does the mispronunciation of *pussy cat* as *cassy put*, wherein the melodic content of the two head nuclei and their licensed onsets interchange. Other logical possibilities, such as [kisæ pʊt] for *pussy cat*, are not attested. Levelt (1992: 9ff.)argues that these facts give credence to the characterisation of syllabification as an independent process whereby (in our terms) melody is accommodated 'late' within an extant prosodic structure: 'Probably the most fundamental insight from modern speech error research is that a word's skeleton or frame and its segmental content are independently generated' (1992: 10). Levelt does acknowledge that this neat modular account has been challenged, since there is evidence for a certain amount of lexical interference in 'phonological encoding': selection errors and encoding errors are not entirely independent: for instance, real words (as opposed to nonsense) are created by encoding errors more often than predicted by chance (Stemberger 1983, and others). But any misgivings about this type of evidence being indicative of syllabification as a no-go area for lexical interference in no way undermine its usefulness in underlining the essentially independent nature of prosodic and melodic representations within the phonological component.

A second example of such independence can be derived from a diachronic change in English. A productive transition in certain realisations has taken place over time, and this is exemplified in (1), adopted from Harris (1994: 35):

(1)

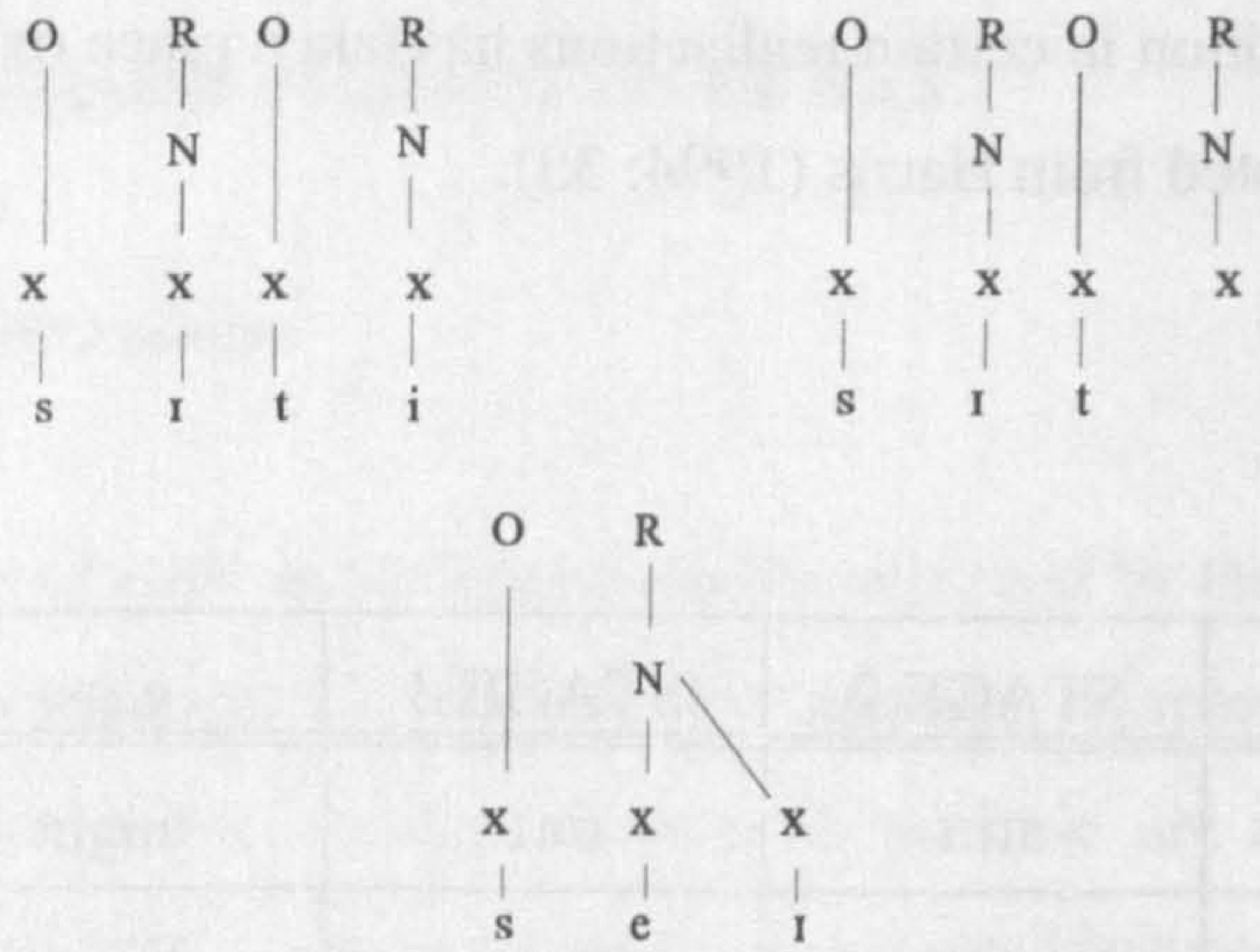
STAGE 1	STAGE 2	STAGE 3	*gloss*
nɪxt	niːt	naɪt	'night'
rɪxt	riːt	raɪt	'right'
wext	weːt	weɪt	'weight'

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 Clements & Keyser (1983), 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 involve 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. One of several prosodic minimality constraints put forward in McCarthy & Prince (1986) states that all English monopod words must be 'bimoraic': in terms of the syllabic model used here, must contain two nuclear positions. Thus different phonetic shapes may be analysed as possessed of structures which are invariant in this parameter. This notion is illustrated for the English words 'city', 'sit', and 'say' in (2)[11].

(2)



---

[11] The empty final nucleus in the word 'sit' has both theory-internal and empirical justifications. Among the first category, the onset-licensing principle demands the presence of this nucleus to license the domain-final onset, and among the second are the vowel/zero alternations to be regularly found in English and other languages (Harris 1994: 179-193).

These three analyses, motivated variously by speech error phenomena, diachronic change and distributional fact, all have in common the demand that we recognise the discrete nature of prosody and melody in the mature system.

This in turn reinforces the conviction that these systems must undergo separate developments. 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' (op. cit.: 94). In our terms, the first of these is akin to prosodic structure, and the second to melodic colour.

Returning to infant perception, what we suggest is that NLMs are evidence of the development of raw melodic contrasts, unattached to the prosodic hierarchy, which itself is developing independently from the perceptual clues to which Jusczyk, Mehler and others have discerned a sensitivity. Corroborative evidence that this assertion is well-founded is gained from the knowledge that humans begin language acquisition pre-natally. The removal from a neonate's perception of any frequency above 400Hz[12] (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison 1988) returns them acoustically to the womb, and the demonstration of their discernment of prosodic cues under these circumstances serves to underline the fact that there must be a chronological mismatch between the beginning of the acquisition of the two systems: we cannot begin to attune to the melodic repertoire of our own language until we get the input, i.e. post-natally, by

---

[12] Referred to in §2.2.0, (p.40) above.

which time our sensitivity to prosody has already been partially reorganised.

If the independent and parallel development of prosody and melody were convincingly demonstrated, we would have the answer to the apparent anomalies between Kuhl and Jusczyk's findings. A clue to another ostensible irregularity in the acquisition of phonology may also be provided by this notion. It is commonplace that at a later stage of life child language production is characterised by gross reductions and neutralisations of the adult repertoire (see, for instance, Smith 1973). There are generalisations to be derived from these phenomena, despite the idiosyncrasies of individual children. How do we account for this apparent loss of facility? Provided that we can satisfactorily align perception with production, the answer to this question follows from the same suggestion. Because we are now proposing a continuum between the development of infant psychoacoustic categories and later melodic representations, these reductions and neutralisations will result from the process of linking up prosody and melody: the departure from the 'state of grace' when these systems are independent would usher in a whole new era of hard labour for the developing psyche.

The attested losses in perceptual finesse after the age of ten months (see §2.2.2 p.46) are accounted for differently by Barrett (1997). Here, the focus is not on the role of phonology, but on a proposed change in the *modus operandi* of prototypes, and a refocusing of (non-modular) attention. Once the child begins to learn its 'first words', attention is 'diverted from (the) sounds of speech' (Barett 1998: 7) and the prototypes function as 'attractors', just as they do in adult speech. By contrast, during the initial pre-word period when prototypes are being developed, they function as 'repellors', enabling fine tuning on particular areas of perceptual space. Barrett likens this early-life ability to her own demonstration of comparatively fine discrimination of prototypical chordal triads by professional musicians (op. cit.: 6). She also states that generalised training (for speech or music perception) works for everyone to fine-tune abilities. A potential problem with this proposed parallel is that the instances of generalised training for specialised perception which Barrett cites are activities which involve a certain amount of conscious attention and the possession of an adult (or at least not an infant) mind/brain.

A question which Barett understandably does not address, since it is not in any way central to her work, is where phonology comes from and when this happens. It is presumably fair to say that at the critical age she has identified, when the child is gaining its 'first words', it must be in possession of some language-specific phonology, and however the change in psychoacoustic perception at that age may be modelled, the said change must map onto a developing linguistic sound-system.[13] This phonology is, in Barrett's terms, part of the 'higher-order processing' whereby the 'attention of the infant is distracted' (op. cit.: 7). But if phonology does represent a different order of processing, by dint of the child's ability to process any phonology at all, it must already be in possession of the abstact categories needed to do so. For vowels, at least, these categories have been nativised by some reference to NLMs, so it is hard to see why they may subsequently become a source of distraction. The *new* work that the infant must do may be rather to align the two essentially separate modules within its phonology.

Another of the critical-period diversions that Barrett suggests that infants are prey to is 'attending to the rhythmic structure of the language' (op. cit.: 7). We have previously in this chapter cited Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison (1988), who show that very much younger children (four days old) attend to some aspects of their native-language rhythms and Jusczyk, Cutler & Redanz (1993), who discerned sensitivity to the trochee in nine-month-old English acquirers. So if anything about language-rhythm is indeed distracting from language-sounds at this subsequent stage, it must be something other than basic stress and intonation formats. Once again, Barrett's proposals are not inconsistent with our move to characterise an intra-phonological 'link-up'.

A further prediction made by the proposal that infant and adult representations are directly related springs to mind, and may bear empirical testing: some prototypical magnets should be stronger than others. We may be able to demonstrate that those which relate to cross-linguistically less marked, more robust, targets develop earlier or more easily. In addition, it may be that 'uncoloured', or inactive, phonological objects don't correlate with an early melodic category at all but are simply expressions of an otherwise unspecified constituent

---

[13] 'Linguistic' is here used in its weaker generic sense: 'pertaining to Language' (see §1.3.4 p.32).

in a particular prosodic relationship.[14]

It may be worth emphasising again that certain traditional constituents and terminologies have no place in the phonological theory we adopt here, and such terms may be an active hindrance in this discussion. The phoneme long ago fell apart into a set of non-linear primal abstracts, whatever theory we subscribe to, and remains only as a shorthand term for a melodic construct. Its definition, though, appears to be one reason why some researchers demur at ascribing any phonological relevance to infant perception. If we persist in according psychological reality to the traditional notion of the phoneme as 'a minimal unit of sound needed to contrast meaning' in a given language, we are bound to develop a confused idea of what exactly an infant is acquiring. If there are no phonemes, then we can't acquire them. This point may seem to some theoreticians an obvious one, but even comparatively recent acquisitional literature is riddled with references to 'phonemic acquisition', and in general texts on speech perception there is a similar conservatism: after discussing contrastive process models, one 1992 text summarises thus: 'Human speech processing is a rapidly paced phenomenon which requires the listener to impose a phonemic identity on incoming sounds...' (Kent & Read 1992).

To help characterise an autonomous phonological level without phonemes in psychological terms it may be useful to countenance a brief diversion while we consider a comparison with visual perception. Waking from a deep sleep, and yet still not fully awake, it is possible to experience visual input without certain top-down conceptual influences. Colour perception is normal, as is a sense of the shape and probable constitution of the contents of the visual field; however, there is neither depth perception nor orientation (no up or down), and without this information the brain is unable to decode the input in terms of *meaning*. Usually, this state of consciousness is replaced after a second or two by normal wakefulness, at which point the objects within the field suddenly flip into their usual relationships, and only then for the first time does the sleeper

---

[14] See §6.3.2 for some discussion of the representation of 'inert' phonological objects, and their possible interaction with the prosodic hierarchy.

have any idea where he is, even if the environment is a very familiar one.[15] Quite how to fit this 'semiconscious' visual perception into a formal cognitive account of vision is not crucial to this discussion, but in some respect the visual field stripped of depth and orientation is highly reminiscent of David Marr's '2½D sketch' (see, for example, Marr 1982).

The relevance of this experience to the perception of speech is that there is a level at which a pattern recognition system operates independently of top-down influences. Neural input is decoded into familiar shapes and colours, but that is as far as it goes. A phonological level has an exactly analogous series of primes, and at this level the multidimensional psychoacoustic space is interpreted by a phonological pattern-recognition system in terms of these primes.

2.4 A direction for empirical research

What kind of empirical evidence would support the idea that infant perceptual abilities bear directly upon the target phonological system? Two propositions spring to mind, and these are set out in (3).

(3)

(a) If there are parallels in the acquisition of two classes that are phonetically dissimilar but are successfully analysed as reflexes of the same phonological prime, this would constitute positive evidence.

(b) As long as the effects of different perceptual strategies are eliminated, there is no reason why one phonetic class should be acquired later than another. However, if a time lag in acquisition is discerned between two classes which use the same mode of perception where the first to be

---

[15] Other people concur that this state of consciousness is familiar: it does not appear to be a personal pathology.

acquired is predicted as phonologically less complex than the second, this would constitute positive evidence for the validity of the phonological model employed here.

In the remaining chapters, most effort will be focused on making some headway in investigating the first of these proposals, the issue of phonetic dissimilarity and phonological identity.

One way of approaching this is suggested by a brief review of 'voicing'. The traditional fortis/lenis bifurcation of English stop consonants is characterised in the theoretical terms used here as a distinction between segments containing the laryngeal element [H] (the aspirated series) and one which lacks this element (the non-aspirated series). 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, by contrast, 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).[16]

The use of any two unary elements delivers a potential four-way contrast. There are languages such as Gujarati and Panjabi in which a four way contrast does, in fact, obtain for phonation (for Gujarati, see Ladefoged & Maddieson (1996: 315); for British Panjabi, the phonetic analysis advanced by Stuart-Smith (1996)). 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: 135).

Now laryngeal elements [L] and [H] are in addition proposed to be potentially componential in nuclear positions where they are the active tonal elements in lexical tone languages (Kaye, Lowenstamm & Vergnaud 1990). There are diachronic arguments for this identification: Hyman (1975) reports that what was once a fully-voiced obstruent in

---

[16] See also §5.2.2, (p.134).

some southeast Asian languages has become at a later stage realised as a low tone on the following vowel. Parallel historical developments are traceable in many languages from this part of the world. Blumstein (1991: 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 similar 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; and finally, a six-tone system without either initial phonation contrasts or final consonants. These stages are exemplified in (3):

(3)

Stage I	pa ba	pah bah	paʔ baʔ
Stage II	pa ba	paˆ baˆ	paˇ baˇ
Stage III	pá pà	páˆ pàˆ	pá˜ pàˇ

Of course such transcriptions are open to all kinds of interpretation, which can give rise to various different phonological analyses. But with that *caveat* in mind, we may be forgiven for speculating that one partial phonological analysis that is at least as good as any other is that the progression between stages II and III represents the transference of the phonation contrast earlier held by the initial consonant to the following (licensing) nuclear position. (The orthography used here may be slightly misleading: note that each stage represents a substitution, rather than an elision.)

A potential drawback of this account of [H]/[L] componentiality is that we have to accept that in non-tonal languages with no fully-voiced obstruents (like most dialects of English), [L] doesn't occur at all, which is an unattractive idea in a theory which aspires to be both universal and economical. Work is in progress that could satisfactorily demonstrate that 'low-tone' and 'nasal' are realisations of the same element. This research is referenced and discussed in chapter 5: it may, in time to come, overcome this particular objection.

Asymmetry also exists in a typological skew found in African languages generally. In these languages, [H] is more generally found as an active tonal component of a nuclear position than is [L]. In Bantu languages, [H] is the only underlying tone, its robustness being attested to by, among other things, the processes of acquisition of Sesotho verb-roots set out in Demuth (1993): children will in this context use a default representation in which [H] is componential rather than a less complex toneless root. This asymmetry needs a formal account and such an account is offered in chapter 5 in the context of west African languages.

Given the 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 the hypothesis that perception at 6-8 months of age is already being transduced into phonological information having systemic import.

If we do ascribe this significance to early perception, it would be mistaken to regard the Spanish/English results reported in §2.2.1 (above) as indicative of a developmental asymmetry such that [H] discrimination is present regardless of target language, and that [L] discrimination only arises if reinforced. The situation is rather that the psychoacoustic discrimination of [p^{h}a]/[pa] is endowed to some mammals,[17] including humans. The perception of the English contrast by the Spanish children in Eilers, Gavin & Wilson (1979)[18] is not best characterised as 'phonetic', but rather as a psychoacoustic hangover.

2.5 Summary

This chapter has reviewed some important work which demonstrates finesse of perception in the first year of life, and specifically shows that language-specific modulation of the

---

[17] And possibly, for all we know, to other organisms.

[18] See §2.2.1, (p.41).

given repertoire of perceptual abilities has taken place by the age of six months. The phonological theory adopted in this thesis, particularly with regard to its central tenet of bi-modularity, allows the proposition that there is a continuum in development between infant skills and adult phonology.

The main stated instrumental aim of the present discussion is to investigate the putative parallel acquisition of identical phonological melodic objects under diverse phonetic guises (§1.2, p. 19). Given the extreme disparity of the nuclear and prenuclear cues for the two laryngeal elements, extant findings on stop contrast perception will therefore be aligned in later chapters with new experiments using subjects who are acquiring a lexical tone language. For reasons to be outlined later (§4.2), Yorùbá turns out to be entirely suited to this purpose.

Before turning to this question, though, the next chapter presents some original evidence for the existence of psychoacoustic prototypes. Phonological theory is used to frame the question, and the results raise clear possibilities for acquisitional testing.

PAGE
MISSING
IN
ORIGINAL

Chapter 3

Language-Specific Prototypes in Catalan

3.1 Introduction

The crux of this chapter is a set of results from original perception tests concerning vowel height. As we have previously remarked, vocalic perception, unlike that of obstruents, has been characterised as prototypical (Kuhl 1992 etc.). Anyone who has attained the age of six months or so is said to have had sufficient language-specific input to warp given psychoacoustic perception so that prototypes for their native vowels exist in vocalic space. Once these begin to be developed, later instantiations of a vowel with similar values to a prototype will be classified as belonging in the same set, so the net result on perception will be a loss of acoustic finesse close to the prototype.

The tests reported in this chapter have been framed to answer phonological questions and could ultimately become relevant to one of the two acquisitional predictions proposed in §2.4 (p.57). This prediction is presented as a logical consequence of the theory of melodic elements (outlined in §1.4, p.33), and concerns a putative link between chronology and phonological complexity. It may be summarised thus:

> The development of different perceptual routines (prototypical, categorical, ...) may certainly be independent of each other in ontogeny. For two 'phonetic classes'[1] which use the same mode of perception, however, a time lag in acquisition is arguably the result of something independent of perception. If such a time lag be discerned between two classes where the first to be acquired is predicted as phonologically less complex than the second, this would be consistent

---

[1] i.e., possessed of two observably different specifications, in conformity to the definition offered in §1.3.3 (p.27).

with a phonological explanation.[2]

For reasons given in §1.5 (p.37) and §3.5 (p.83), a final conclusion on this question will have to be postponed beyond the remit of the present thesis. Instead, in the succeeding chapters we will turn our attention to the matter of the parallel acquisition of diverse melodic objects exhibiting phonological identity. Nevertheless, the work reported here fulfils the following objectives:

(i) It serves as a preliminary operational exercise for the tests to be reported in chapter 4.

(ii) The language-specificity of vowel prototypes is clearly displayed, thus corroborating Kuhl's NLM theory[3] from independent evidence.

(iii) The results allow us to frame a clear research objective for future acquisitional research into the chronology/complexity issue.

In the next section we describe the suitability of Catalan for our purposes. This is followed by accounts of the synthesis and verification of appropriate tokens for testing (§3.3), results of adult perception tests which deliver supportive evidence for Kuhl's earlier findings (§3.4), and the delimitation of a research project concerning infant perception (§3.5). Finally, we summarise in §3.6.

3.2 The vowels of Catalan

In recent times there has been some neutralisation of the mid-vowel height contrast in

---

[2] The relatively late acquisition of complex *prosodic* structures, such as branching onsets, by older children has often been remarked (see, for example, Smith 1973). The search for a link between chronology and complexity in melody is encouraged by this parallel.

[3] See §2.4 (p.57).

Catalan, owing to the geographical proximity of speakers of Castilian Spanish, who do not natively maintain this contrast. Despite this, however, both the front and back mid-vowel distinctions remain robust for many Catalunyans, especially 'up country' of Barcelona. We therefore derive, for this language, an inventory of seven peripheral vowels plus one neutral vowel (which is phonetically lower than most varieties of English schwa). The vowels of Catalan are arrayed in (1).

(1)

i u
e o
ɛ ɐ ɔ
a

The neutral vowel [ɐ] only occurs in weak prosodic positions and as an alternant of any of the set of vowels [e ɛ a] (Palmada 1991). In the model adopted here, reduction processes generally consist in the stripping away of elemental material, so the neutral vowel can be regarded as a realisation of any of the three lexically-present vowels with elements [A] and [I] supressed (Harris 1994: 113). The alternation is illustrated by the suffixation of [pa] 'bread' with [+net] to produce [pɐ'nɛt] 'small loaf'.

Notwithstanding the recent neutralisations referred to above, the seven peripheral vowels can be fully contrastive for four heights, as witnessed by the list of minimal pairs in (2).

(2)

si	*yes*	nu	*naked*
se	*I know*	no	*no*
te	*here you are*	do	*note (mus.)*
tɛ	*tea*	dɔ	*gift (of character)*
kɛ	*what?*	bɔ	*good*
ka	*dog*	ba	*he goes*

Catalan has the fortunate quality (for phonologists) of bearing contrasts on a single parameter, rendering it easily accessible as a target for testing.

Height contrast amongst mid-vowels, as represented by the two pairs [te tɛ] and [do dɔ] in the Catalan examples, is typologically marked. An asymmetry is attested: Maddieson's (1984) list of 317 languages contains 78 (25%) with a five vowel system of the /i e a o u/ type and a further 24 (8%) using this inventory plus a central vowel of some type, while only 19 (6%) are listed which utilise the seven vowel system, with a further 3 (1%) adding a central vowel. Furthermore, there is a distributional asymmetry to be found in a system *without* such a contrast. For *front* mid vowels, the lower front vowel is more often present. In those five vowel languages listed in Maddieson where the height of the front mid vowel is further specified, the lower is identified in Greek, Zulu, Burera (an Austronesian language) and nine other languages, whereas the higher appears in only two: Luvale (a Bantu language) and Garo (Sino-Tibetan). The opposite asymmetry appears to hold for back mid vowels. It may be possible to discern parallels between these facts of typology and the phonological theory outlined in §1.4 (p.33). This possibility is aired in §3.5, but for now attention is going to be concentrated on an operational effort to gain some confirmation of the existence of language-specific prototypes in vocalic space.

3.3 Creation and verification of stimuli

Natural tokens of the minimally contrastive pairs which utilise front vowels identified in (2) above ([si se; te tɛ; kɛ ka]) were produced by a native speaker of Catalan (who also speaks English). These 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.

Five tokens of each word were recorded and analysed using a set of programs which are generically termed *sfs*[4]. These allow, among other things, audio reproduction, together with a visual display of both the amplitude signal and the spectrograph.

The samples were examined to identify typical duration and formant values for the vocalic portion of each word with particular attention being paid to F1, a primary acoustic cue for vowel height. To control for the effect of 'artificial' pronunciation as far as possible under the circumstances, the target vowels were also embedded in two minimally contrastive sentences. Since our interest in [i] and [a] is tangential at present, attention was focussed on the mid vowels and three tokens of each of the utterances in (3) were recorded.

(3)

(a) Tinc els ulls cecs amb tanta llum
[tiŋk ɐls uʎs **seks** ɐm tantɐ ʎum]
I have | the | eyes | blind | with | so much | light
'My eyes are blinded by all this light.'

(b) Tinc els ulls secs amb tanta llum
[tiŋk ɐls uʎs **sɛks** ɐm tantɐ ʎum]
I have | the | eyes | dry | with | so much | light
'My eyes are dry from all this light.'

For the tokens of any particular vowel in the bare citation forms, there was little or no variation in duration. Measuring from the onset of vocal fold vibration to the point where a significant loss of energy is detectable, all the tokens exhibited a value of about 160ms, considerably longer than the sentential vowel tokens, which average under 100ms. This is presumably at least partially due to the 'artificiality' of the pronunciation.[5] Confining our observations still to the isolated utterances, no more than 5 Hz variation was observable in any formant at the onset of vocal fold vibration. F3 for all the words hovered at 3250

---

[4] 'Speech Filing System', developed by Mark Huckvale at UCL. See also §4.2.2 (p.89).

[5] For examples, see Figure 3.1 below.

Hz throughout. A lowering in F3 is associated with rounding, so we may have had to take F3 into account had we been using back vowels as the subject of our investigation, but for present purposes it was subsequently ignored.

In contrast with the lack of intra-category variation for F1 or F2 at onset, the predictable clear inter-category disparity for the value of F1 emerged, together with a regular shift in F2, again predictable from the fact that a higher front vowel is extremely likely to be articulated further to the front of the vocal tract.

The frequency value of the centre of F1 and F2 did, however, vary during the vocalic segment. In order to calculate an average for the purposes of synthesis and subsequent verification and testing, a typical token of each citation word was selected, and seven measurements of the frequency of these formants were taken at aproximately 20ms intervals during the vowel. These measurements were averaged, and the results established that, for this particular speaker pronouncing isolated words, F1 and F2 for the four unrounded unreduced vowels of Catalan have the values given in Table 3.1 (to the nearest 5Hz).

	i	e	ɛ	a
F1 (Hz)	380	420	570	905
F2 (Hz)	2620	2580	2500	1740

Table 3.1: Average F1 and F2 values for one native speaker of Catalan

A comparison with the sentence-embedded vowels produced in the words *cecs* and *secs* (as in (3) above) revealed some disparities, but significant similarities, with these results. Durationally, all the vowels were shorter than in the bare citation forms, an entirely predictable fact. There was also more durational variation, both intra- and inter-category: *secs* displayed, on average, a longer vocalic portion than did *cecs*, as can be seen by comparing the typical spectra in Figure 3.3 (below). Whether or not durational disparities play any part in the cueing of a vowel difference is impossible to say purely from this evidence, but it seems a far-fetched proposal. It is possible that the durational variation in our sentential samples reflects the universal (mechanical) tendency for lower vowels to be longer, while the lack of this disparity in the isolated syllables reflects the 'carefulness' of the pronunciation.

Far more soild ground for the identification of an acoustic cue is presented by the F1 values, which hardly depart from those identified in the citation forms. For [e], values are more or less identical, while for [ɛ], there is a discrepancy of a maximum of 20 Hz. F2 was somewhat lower, on average, in the embedded than in the isolate vowels, indicating the more central articulation of the vowels in running speech. The most significant disparities in this respect were observable in tokens of [e]. Toward the end of the vocalic portion of [seks], F2 does run through its predicted value *en route* to the velar locus frequency (see the arrowed line in Figure 3.3(a) (below), giving the auditory impression (to a non-native speaker of Catalan) of a diphthong, but in the main, it is true that F2 is lower in the embedded than in the isolate examples. F3 follows suit less dramatically, being generally slightly lower than 3250 Hz in the sentential tokens. The more careful, more peripheral, articulations in the bare citation forms bear some kind of inverse relationship to the number of other linguistic cues (of any type) in the utterance. Particularly with regard to the invariance of F1, it is fair to propose that the connected-speech samples provided no evidence to invalidate our carrying out perception testing using single words.

The KPE80a parallel synthesizer[6] 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. This was used to create artificial copies of the six words from (2) above which utilise a front vowel-height contrast. These were given the F1 and F2 values indicated in Table 3.1. The synthesis of the artificial stimuli allows the manipulation of formants independently of any other parameter, so a single master was used to create each minimal pair by exclusively varying values of F1 and F2. Three tokens of each of the six words

---

[6] Developed by Andrew Simpson at UCL. See also §4.2.3 (p.93).

were then randomly ordered, and for verification purposes the resulting list of eighteen words was replayed through headphones to the Catalan informant, whereupon English translations were requested. 100% successful translations were provided for the mid-vowel minimal pair[7], and it was concluded that the artificially produced tokens of this pair bore genuinely representative F1 and F2 values for this speaker.

It is appropriate to mention that, although we are looking for felicitously manageable synthetic speech stimuli for testing, and although there is a direct relationship between vowel height and the value of F1, there is a limit to how simplistic we can make the stimuli, and if we *only* varied F1, we would be overstepping this limit. Perceptual normalisation must take account of the relative position of a vowel within (at least) the space defined by F1 against F2 (or F1 against F2 - F1), otherwise the complete lack of overlap between the F1/F2 values produced by speakers with very different vocal-tract sizes would be incomprehensible. A research project set up at the University of Barcelona which has been studying the putative perception of vowel native-language-magnets (*à la* Kuhl) by speakers of Catalan and Castilian Spanish has found a somewhat different prototype for [e] than the one we identified in Table 3.1 above. Our prototype sits at 420/2580 Hz in vocalic space, while theirs is at 404.8/1929.8 (Bosch, Costa & Sebastián 1994). This is less worrying than it would be if we did not consider the place of normalisation in perception, which must logically bring about some global shift in the physical parameters of an invariant multidimensional psychoacoustic space from one moment to the next. The fine tuning of this process has not yet been described or explained, but the only anomaly between two predicted values for a particular prototype that would ring genuine alarm bells is one which increased one of the parameters while decreasing the other.

Six annotated spectrograms are presented in Figures 3.1, 3.2, and 3.3 (pp. 71-73) to illustrate both the comparison between the isolate and embedded natural tokens of [e] and

---

[7] Perception of the low/ mid-low pair was less successful, presumably because of some flaws in the synthesis -[a] is notoriously hard to artificially produce because the high F1 tends to be incompatible with the reinforcement of the perceptually important second harmonic. This pair, however, was not the focus of subsequent testing so we proceeded without refining these stimuli further.

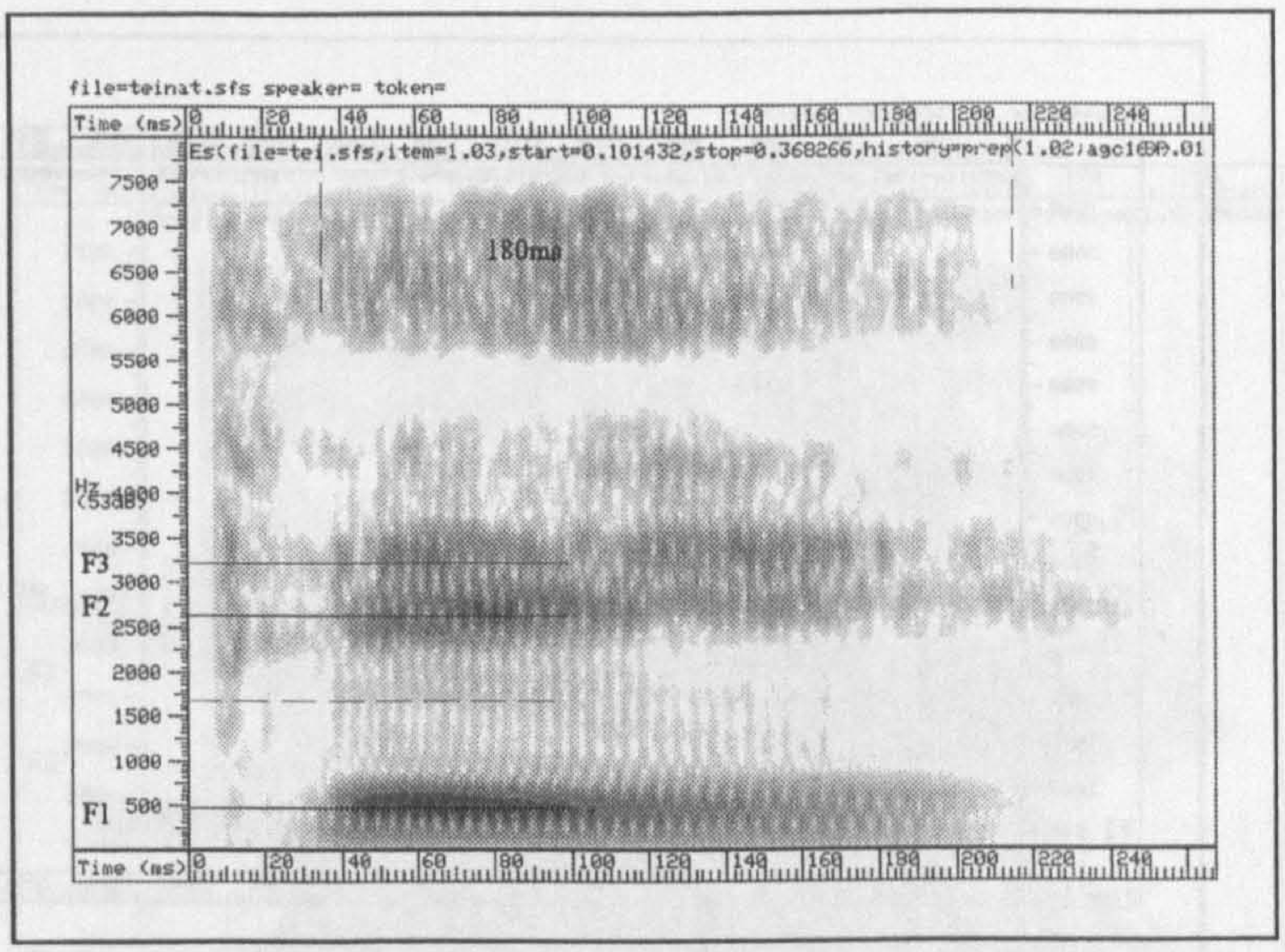


Figure 3.1(a) Natural token of Catalan [te].

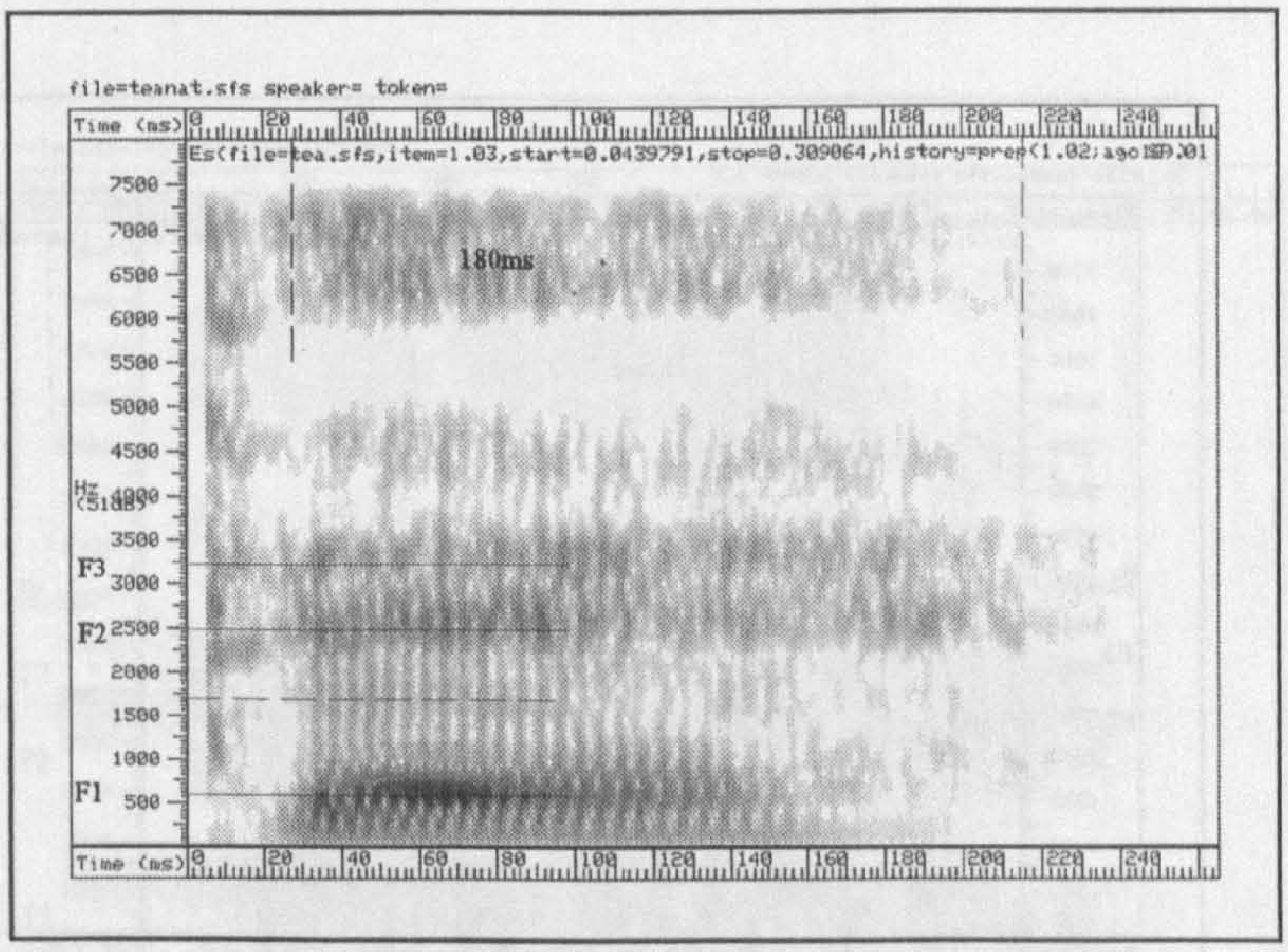


Figure 3.1(b) Natural token of Catalan [tɛ].

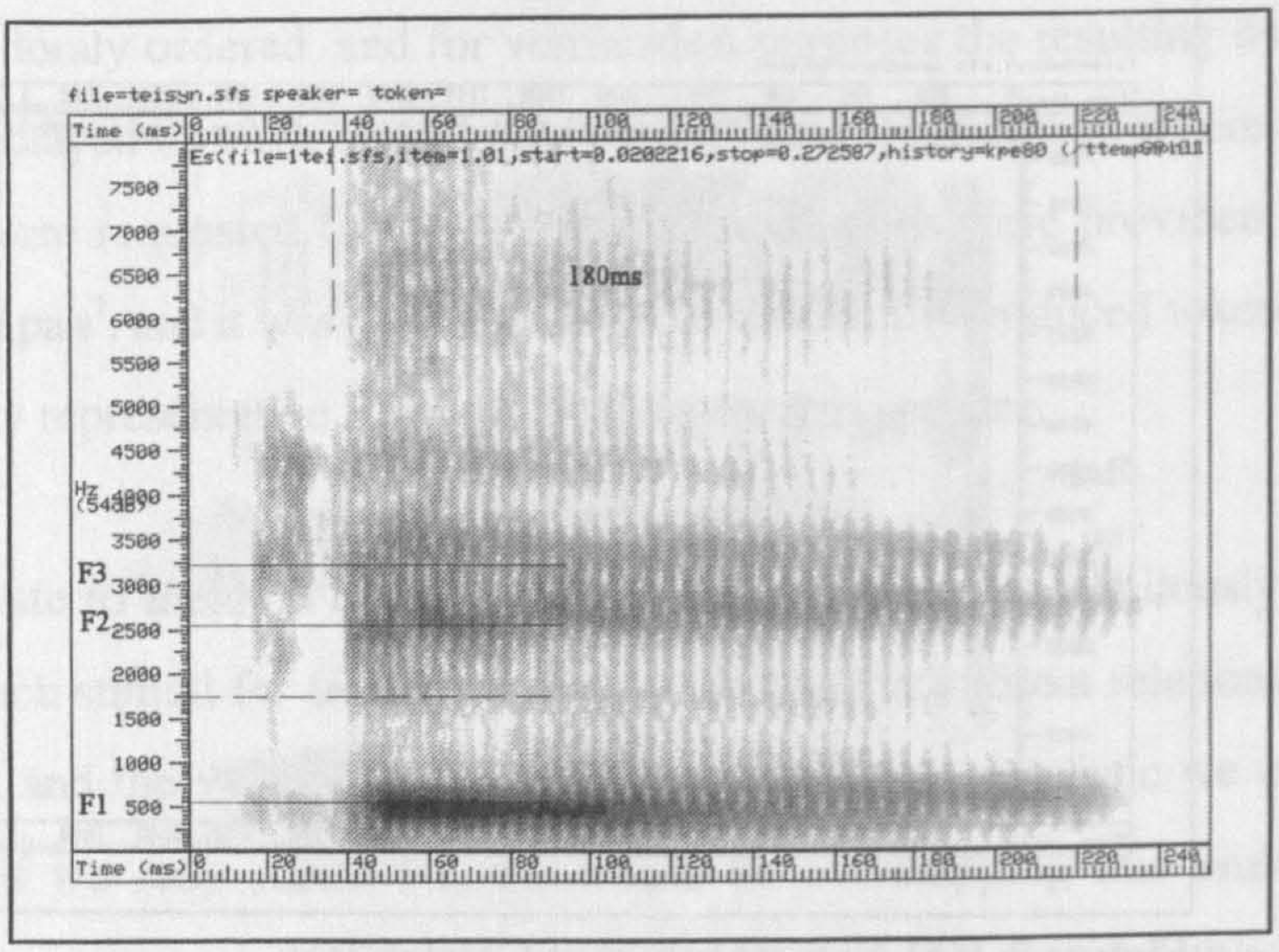


Figure 3.2(a) Synthetic token of Catalan [te].

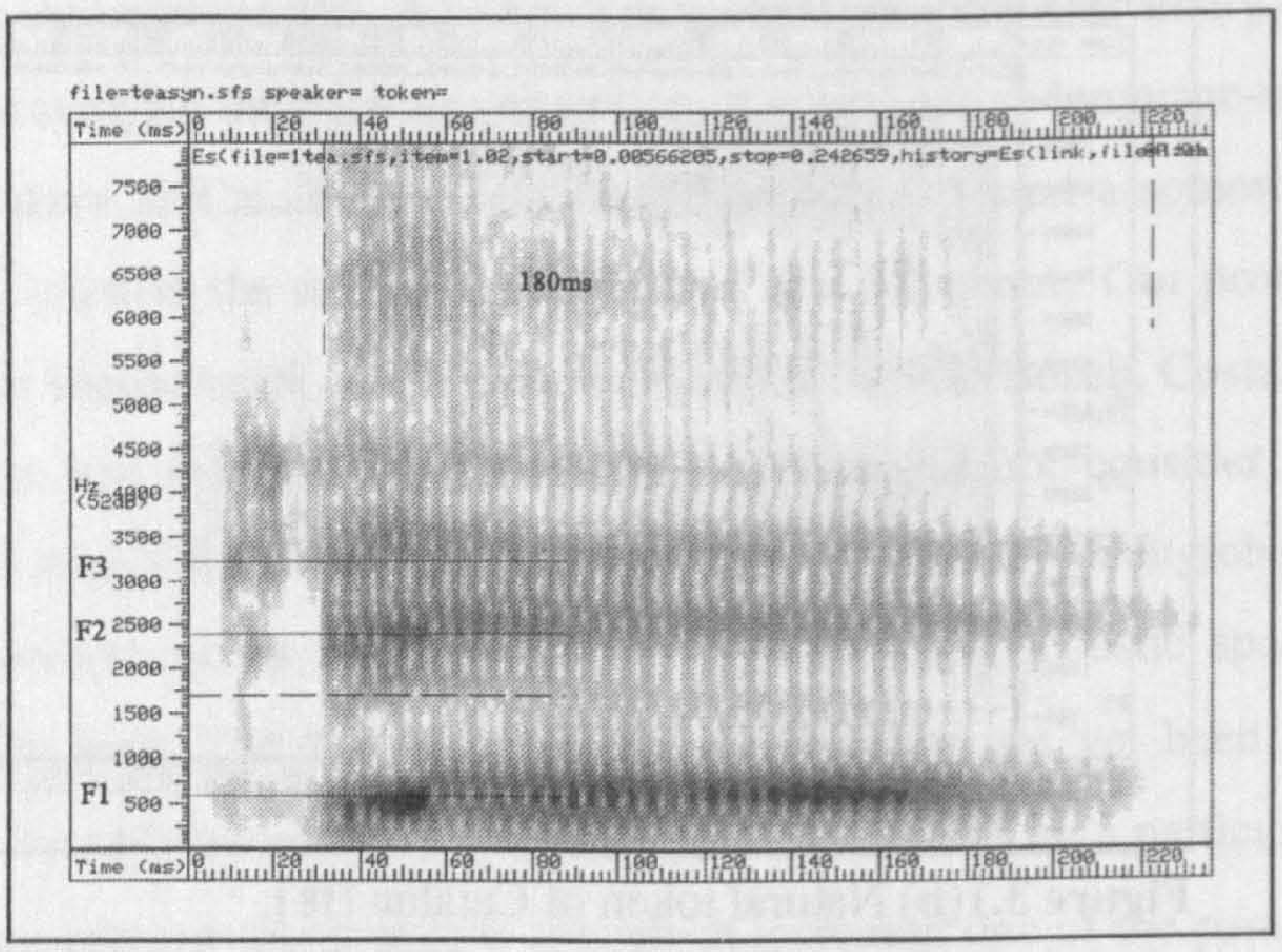


Figure 3.2(b) Synthetic token of Catalan [tɛ].

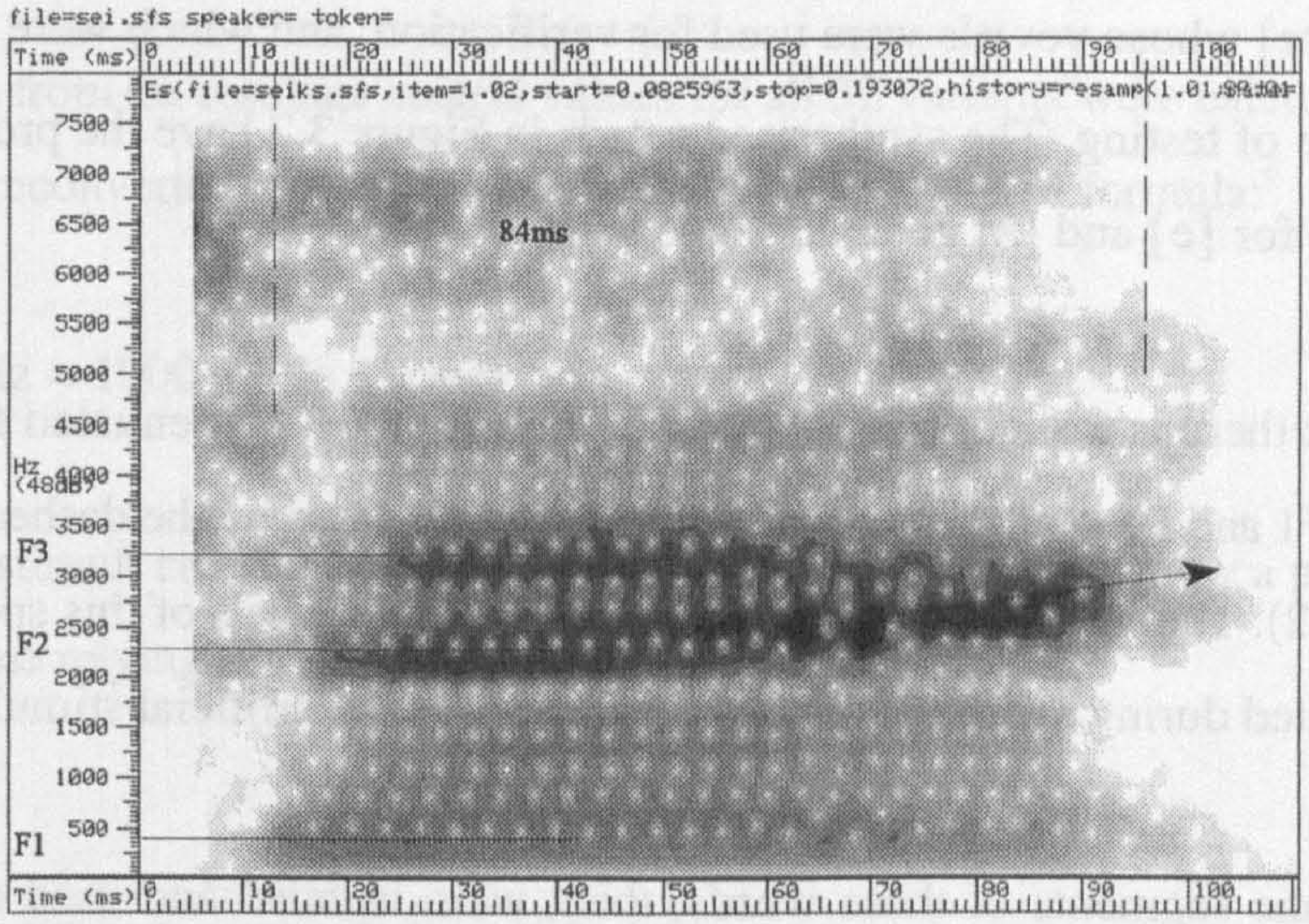


Figure 3.3(a). Sentence-embedded token of Catalan [(s)e(ks)].

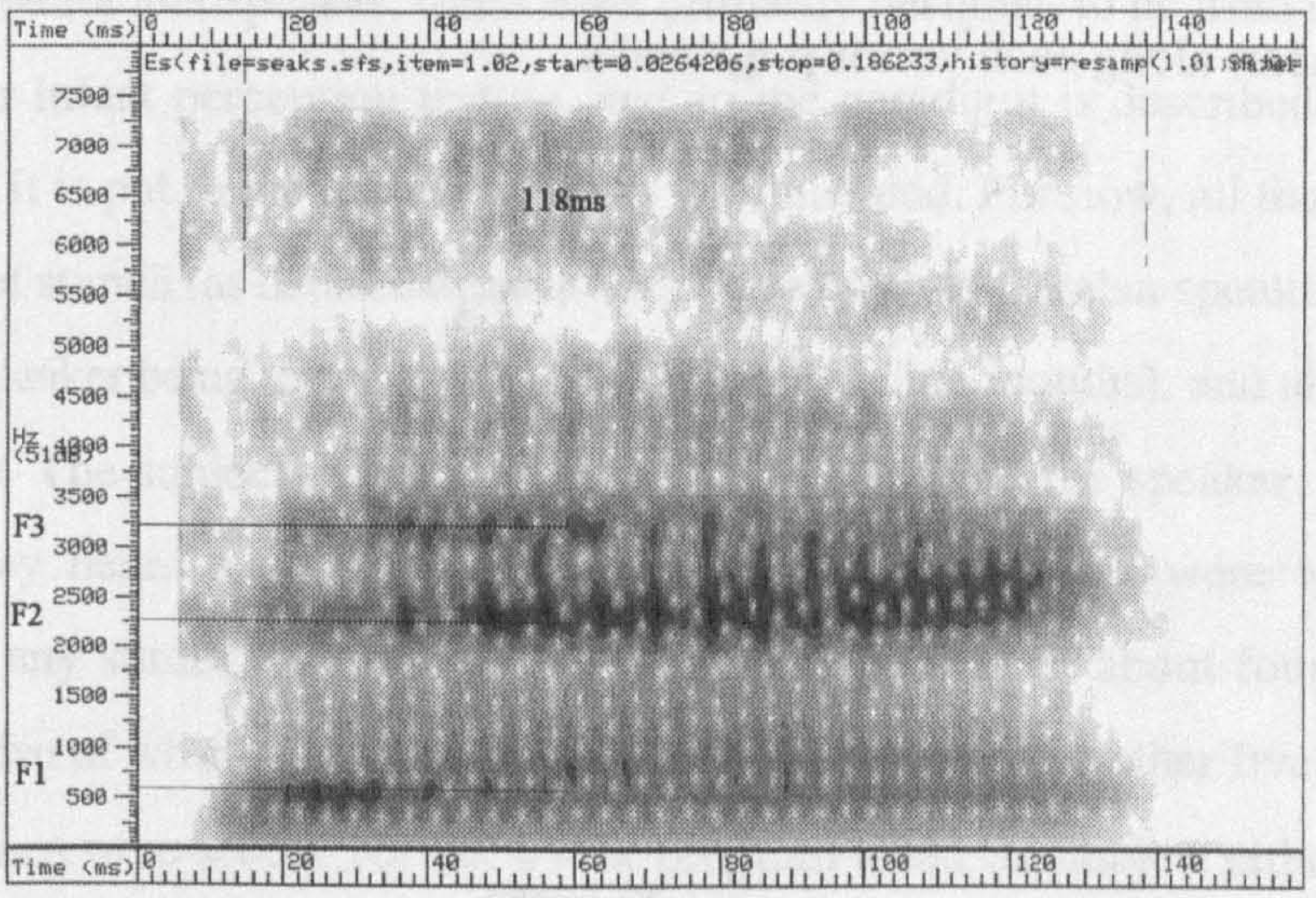


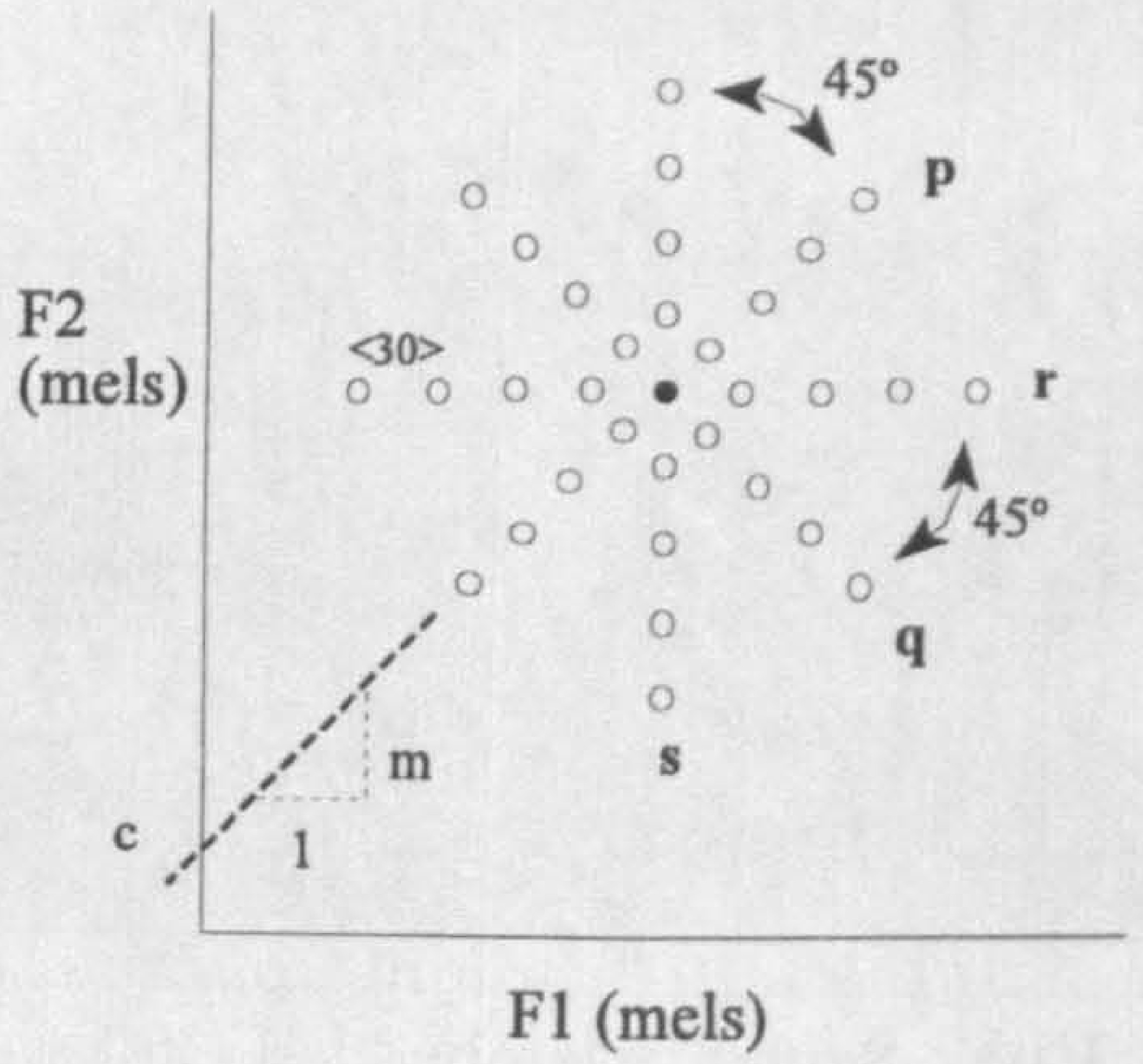
Figure 3.3(b) Sentence-embedded token of Catalan [(s)ɛ(ks)].

[ɛ] referred to above, and also the relationship between these and the synthetic tokens of [te]and [tɛ] whose vowels were used for verification, and which were to be used in the next stage of testing. The synthesised words in Figure 3.2 have the prototypical F1 and F2 values for [e] and [ɛ] given in Table 3.1.

Note that there is a slight intensity peak, like a somewhat attenuated formant, situated between F1 and F2 in all these spectrograms (and indicated by the dashed lines in Figures 3.1 and 3.2). This was a characteristic of all the natural vowels of this speaker, and it was incorporated during synthesis to echo this feature in the artificial stimuli.

The vocalic segments of these words, then, were isolated and used as prototypes. A constellation of perceptually equidistant variants, each variant 30 mels apart, was produced for each prototype, as diagramatically represented in (4), where the central dot in the array represents a prototype and the dots in the outer rings represent variants sited 30 mels apart along each vector **p,q,r** and **s.**

(4)



To achieve this, the Hz values for each prototype were first converted into mels using the formula:

$$\text{mel} = \frac{1000}{\ln(2)} \text{ x } \ln(1 + \frac{\text{Hz}}{1000})^{8}$$

Distances in mels from the resultant central values for all 32 variants were calculated, and all these values reconverted into Hz for synthesis using the reverse formula:[9]

$$\text{Hz} = 1000 \text{ x } (\frac{\ln(2).\text{mel}}{1000} -1)$$

All the resultant stimuli could then be paired in any combination for use by a perception testing programme relying on 'Same/Different' discriminations.

3.4 Perception of Catalan vowels by adults

The *Turner* programme[10] relies on the subject discriminating a change in an acoustic stimulus played over a loudspeaker. It has been primarily designed to be used in tandem with VRISD for infant perception testing, and so the paradigm is described in full in chapter 4, where it is put to the use for which it was intended. For now, all that need be said is that pairs of stimuli (as detailed below) were played to two Catalan speakers during three tests (one speaker being tested twice at an interval of nine months), and also to two English controls. The subjects, who were alone in a room with the speaker, pushed a button when they heard a change in stimulus, and their responses were logged by computer.[11] For any stimulus-pair there were ten *Trial* periods of about four seconds duration, during five of which the stimulus did change, and during the other five of which it did not. Responses were scored 'correct' by the program when the subject either pushed the button when the stimulus changed, or refrained from doing so during a *Trial* when the stimulus remained unchanged.

---

[8] The use of this formula is discussed, among other places, in Fant (1973: 47-48).

[9] For details of these calculations see appendix D.

[10] Developed by Mark Huckvale at UCL.

[11] See §4.3.1, (p.101) for a full description of the experimental paradigm, and appendix B(ii) for a sample 'Turner' log. Reinforcement by pink drumming penguins is of course optional for adults.

The stimuli were separated by a minimum of 30 mels along any of the vectors within one of the prototype-centred arrays. The hypothesis being tested was that familiar from the work of Kuhl *et al.* (1992 etc.), namely that there would be a loss of finesse of perception around a vocalic prototype for a native speaker of a language. Bosch, Costa & Sebastián (1994) have already achieved positive results for the recognition of an [e]-prototype for Catalan, but not so far for an [ɛ]. In the present tests the clearest indications would ideally be obtained from the results drawn from perceptually equidistant stimuli (separated, on the basis of figures taken from previous research, by 30 mels) at different removes from the prototype. It should be noted that stimuli separated by this number of mels are not, on casual listening, easily distinguishable at all, and that this can be regarded as a positive factor in the program. If a statistically significant result is obtained in any of the tests, it appears that we are observing a perceptual process at least partially unmoderated by consciousness.

The results of the tests will first be discussed speaker by speaker, and then any generalisations can be made explicit. Pairs separated by 120 mels were used at the start of testing, or after a pause, in all test runs to accustom the subject to the task and the paradigm, and were invariably perceived correctly. Table 3.2 presents results of the subsequent tests for each prototype on the first of the Catalan speakers.

	Stimulus Pairs (0 = prototype; higher numbers are progressively more distant, spaced at equal 30 mel intervals)			
[e]-type stimuli	0-1	1-2	2-3	3-4
(i) Catalan speaker 1 test 1 (East vector)	✗ ✗	✗	✓	✓ ✓
(ii) Catalan speaker 1 test 2 (Southeast vector)	✗	✓	✓	✓

[ɛ]-type stimuli	0-1	1-2	2-3	3-4
(i) Catalan speaker 1 test 1 (East vector)	✗ ✓	✓	✓	✗ ✓
(ii) Catalan speaker 1 test 2 (Southeast vector)	✗	✓	✓	✓

Table 3.2: Results of perception testing on Catalan speaker 1: for an explanation of the symbols, see text.

Table 3.2, (and tables 3.3, 3.4 and 3.5 below) are to be read as follows. They provide results for pairs of stimuli separated by 30 or 60 mels along any vector. The numbers 0-4 in the top row indicate distance from the prototype in 30 mel steps, thus the first column (0-1) shows results for a prototype paired with a stimulus 30 mels distant, column 1-2 uses stimuli respectively 30 and 60 mels away from the prototype etc. '✓' in a column indicates that the difference was perceived, either 100% correctly or at a statistically significant level. '✗' means no difference was perceived, or else the difference was not perceived at a statistically significant level. 'At a statistically significant level' here means in practice a maximum of one error in ten trials, as this conforms with the proposals of Aslin & Pisoni (1980a) with respect to the use of VRISD by Eilers,Gavin & Wilson (1977).[12] Where two symbols ('✓' or '✗') appear in a cell, the stimulus pair was used for testing twice in the same session, and the first result recorded appears before the second.

At first blush, the results seem to indicate the presence of native-language magnets for these Catalan vowels fairly positively. Two factors, though, need to be borne in mind. Firstly, as a session progresses, an adult subject appears to tune in general perceptual strategies and to become more skilled at this type of discrimination than they were at the start. In the first test for [ɛ], it is apparent that *both* the pairs of stimuli '0-1' and '3-4' were misperceived at first pass, and both perceived correctly second time around. Because of this, the crucial '0-1' pair was subsequently presented at different stages of a particular test run to attempt to minimise this effect. More convincing results could certainly have been obtained by always presenting '0-1' at the start (i.e. cheating). Further to this, it should be noted that the results in the tables in this chapter are not presented in the temporal order

---

[12] See also §4.3.1, (p.104).

in which the testing was carried out. For ease of reading, they instead follow a sequence which moves progressively further out along a given vector.

The second proviso to be aware of when analysing the figures in Table 3.2 is the choice of a particular vector for testing. Row (i) in each case (Catalan native speaker 1 test 1) gives results using the 'East' vectors, while row (ii) (Catalan native speaker 1 test 2) employs 'Southeast'. The results for 'Southeast' looked particularly convincing, so these same stimuli were reused for the majority of tests on the first of the English controls. The results for these tests are given in Table 3.3.

[e]-type stimuli	0-1	1-2	2-3	0-2	2-4
English control 1 (Southeast vector)	✗	✗	✗	✓	✓

[ɛ]-type stimuli	0-1	1-2	2-3	0-2	2-4
English control 1 (Southeast vector)	✗	✗	✗	✓	✓

Table 3.3: Results of perception testing on English control 1

Another factor is now evident. Having perceived the training pair (separated by 120 mels) reliably, once testing began it emerged that this subject only perceived an acoustic difference between pairs that were *more* than 30 (but putatively less than 60) mels apart. Nevertheless, no worsening of perceptual finesse can be observed as a result of the presence of prototypical 'magnets'. A 60 mel differential is as reliably perceived close to the Catalan prototypes as it is at some remove.

The larger perceptual distance (>30, <60 mels) needed to be maintained for the second Catalan speaker to perceive differences, but in this case there is indeed evidence of a warping of perception close to the hypothesised 'magnets': results are given in Table 3.4.

[e]-type stimuli	0-1	2-3	0-2	2-4
Catalan speaker 2 (Southeast vector)	×	×	×	✓

[ɛ]-type stimuli	2-3	0-2	2-4
Catalan speaker 2 (Southeast vector)	×[13] ×	×	✓

Table 3.4: Results of perception testing on Catalan speaker 2

After the usual accustomisation (using a 120 mels differential), it became clear that this subject failed to perceive the 30 mel difference. The two failures logged for 2-3 ([ɛ]type stimuli) are taken from two different vectors, and helped to lead to this conclusion. The assertion that there is a warping of vowel space is therefore made on the basis of the contrastive results achieved for the 0-2 and 2-4 stimuli.

Finally, the second English control provided results which indicate a finer acoustic perception than the first control, but also show that this perception is unaffected by the hypothesised Catalan prototypes. Results are given in Table 3.5.

[e]-type stimuli	0-1	2-3
English control 2 (Northwest vector)	✓	✓

[ɛ]-type stimuli	0-1	2-3
English control 2 (Northwest vector)	✓	✓

Table 3.5: Results of perception testing on English control 2

In sum, these results do support the hypothesis that psychoacoustic perception is warped by native-language prototypes. Both prototypes appear to have equal perceptual status

---

[13] The first of these two '×' results was obtained using stimuli from Northwest vector.

for adult native speakers, so there is no obvious correlation between perception in the steady-state language and the typological asymmetry discussed in §3.2 (p.66). The results also indicate that testing of this type has to control for individual variability in acoustic perceptual finesse, a finding not previously reported.

Neither have we yet considered the possibility of 'interference' from a prototype being present for the English front-mid vowel(s), in the case of either the controls or the Catalan speakers, who are both fluent English speakers. It follows from the theoretical proposals being tested here that a prototype or prototypes must exist, at least for the monolingual controls. However, the present results indicate that any such prototype must be sited elsewhere in vocalic space. This presents another potentially fruitful line of enquiry for another day.

3.5 Phonological motivation for infant perception tests

Armed with our demonstration of the presence of a prototype for each of the front-mid Catalan vowels in the steady-state language, the next step could be to see if such items are discernable in infancy. A pilot study has been carried out at the University of Barcelona, using three babies aged around six months from monoglot Catalan families.[14] Unfortunately, none of the children provided us with assessable data, for reasons independent of the testing. One of the children did appear to train successfully on stimuli of 120 mels differential, but then proved restive when testing began.

A further motivation for pursuing this line of enquiry which has not yet been detailed requires a discussion of what exactly is meant in this context by 'phonological complexity'. The adult tests show a straightforward link between perceptual routines and two discrete melodic objects. The reason for hypothesising that one of these may be earlier-acquired than the other, though circumstantially supported by typology (§3.2, p.66), is essentially purely phonological.

---

[14] Many thanks to Dr. Laura Bosch for her help in procuring these subjects.

The front mid vowel contrast is phonologically asymmetric. In an orthodox elemental account, [ɛ] (the lower vowel) is unheaded and is componentially simply [I, A] (unordered), while [e] is [I]-headed ([A I]). Depending on who you read, headedness in Government Phonology (henceforth GP) can be either the property of an individual element (Kaye, Lowenstamm & Vergnaud 1985)[15] or of the tier upon which an element resides (Rennison 1990).

A logical objection to this bald statement of asymmetry may be raised if we query its validity in the light of competition from a symmetrical account where [I] heads [e] and [A] heads [ɛ]. However, GP analyses are consistent in explicitly stating the complementary behaviour of [A], on the one hand, and the 'colour' elements [I] and [U] on the other, with respect to headedness. The 'natural lexical headship' which inheres in the two elements underlying high vowels ('[I] and [U] must be heads') is in contrast to [A] *au naturel*, which tends by contrast to conform to the constraint '[A] must not be a head' (Charette & Göksel 1998, Cobb 1995, and plainly set out in Walker 1995: 110). This kind of statement is generally employed to reduce the generative power of the GP model, and in the absence of further constraints, these two particular statements deliver as an unmarked system the asymmetrical rather than the symmetrical analysis. Empirical support for this, albeit by analogy, can be derived from a language such as Danish, where four front vowels contrast in height only (each can also be long or short). Examples to illustrate this are given in (5) (taken from Ladefoged & Maddieson (1996: 289)):

(5)

vi:ðə	*know*
ve:ðə	*wheat*
vɛ:ðə	*wet* (vb.)
væ:ðə	*wade*

A contrast between [æ] and [ɑ] is also observable in the language, even though there are not a vast number of minimal pairs. Those that do exist include the words [snæk] 'snack'

---

[15] This is also the position taken in Dependency Phonology (Anderson & Jones 1974).

and [snɑk] 'small conversation'.[16] Using isomeric relationships contracted between [I] and [A], the only *logically* possible account of these distinctions is that the lowest of the four front vowels ([æ]) is [A]-headed (a marked structure, as we have said, for GP phonologists), the high mid [e] is [I]-headed, and [ɛ] is an equal-status combination.

Returning to Catalan, the ascription of headship to the higher of the the two front mid vowels now presents the language as an exemplar of the unmarked case with respect to the distribution of headed/unheaded complexes and the related constraints: it has an [I]-headed and an unheaded expression, but no [A]-headed expression.

Within GP, there are other tales of componentiality which do not invoke isomerism at all. The enterprise detailed in Backley & Takahashi (1998) entails the activation of universally-present complements to define the class-property of the ATR set of vowels. For the vowel-harmony processes on which they are focussing their attentions, they argue that complement-activation, rather than headedness, subtends this process. In this system, *all* phonological operations are to be accounted for as the activation *versus* non-activation of primes. For our present purposes, we should note that this competing account aligns with its rival insofar as there needs to be more phonological 'activity' in a high front mid-vowel than in a low front mid-vowel. Indeed, even if we backtrack to more conventional feature matrices, as long as we rely on markedness theory to avoid the worst profligacies of overgeneration, the privative feature [ATR] would be said to inhere in the higher vowel, but the lower would simply lack this specification.

It is explicit in the feature-based approach, and also in the 'complement-activation' account, that greater complexity is present in the higher vowel. But is there any necessary relation between *headedness* and complexity, or is isomerism purely relational? If we were to propose that headedness *were* a source of phonological complexity, the discussion so far would align well with the distributions within five-vowel languages to which we already referred: unheaded ([ɛ]) expressions are the norm. But it should be noted that the

---

[16] Thanks to Pierre Millinge for these examples.

neatness of this parallel between typology and licensing constraints would be more problematic were we discussing back vowels, where the related prediction of a wider distribution of [ɔ] (lower) expressions is not borne out. Even if we do not propose a link between headedness and complexity, the fact remains that typologically [I]-headed objects are more pervasive than [U]-headed objects in this respect. There are distributional asymmetries relating to individual elements which at some stage require further explanation, a point to which we return in chapter 6 (§.6.3).

Given the foregoing, the front-mid-vowel height contrast in Catalan, uncluttered by considerations of vowel length, nasalisation or anything else, bears a potential seam of linguistic data which can be mined to pursue an empirical enquiry into phonological complexity. Research already reviewed (Kuhl 1992 etc.) suggests that at least some prototypes for vowels are acquired by six months of age, so it may be possible to ascertain if phonological asymmetry is reflected in this case by a delay in the acquisition of the more complex object.

It is obvious that we are going to need a large number of children, at different stages of development, to properly test this hypothesised link between chronology and complexity. Furthermore, there are limitations and potential errors of the VRISD paradigm which have not yet been broached. These became evident while testing children for the programme to be described in chapter 4, so a discussion of these issues is postponed until those tests have been described.[17] Pursuing this course of enquiry using a behavioural task alone may not prove to be the most efficient way to proceed. Developments in non-invasive methods of measuring brain activity directly may lead to less problematic paradigms than VRISD, and given some improvements in the ergonomic design of these mensuration techniques, they may eventually be beneficiently harnessed in tandem with behavioural procedures.

---

[17] See §4.4, (p.117).

3.6 Summary

The results presented in this chapter confirm the existence of NLMs for Catalan front mid vowels in the steady-state language. Unlike the English controls, Catalan speakers exhibit a loss of finesse of perception close to postnormalised values identified for [e] and [ɛ]. This finding aligns generally with Kuhl's cited work, which uses parameters other than vowel height. It also partially concurs with the experiments reported in Bosch, Costa & Sebastián (1994). In this last case, as long as the role of normalisation is acknowledged, both results may be seen to agree on the existence of an NLM for the front high-mid vowel. The present finding with respect to the lower vowel does not tally with the Barcelona results, and further confirmation would be desirable.

These results can potentially be linked with the acquisition of phonological melodic expressions of differing compexity. However, any investigation into a connection between acquisitional chronology and complexity is going to require a large number of infant tests, using children of varying ages. For this reason, the matter will not be taken further here. Instead, we turn our attention to a more manageable instrumental project: the acquisition of phonological objects displaying disparate phonetic identities.

Chapter 4

Experiments with Tone

4.1 Introduction

In this chapter, some experimental confirmation is sought for the theoretical proposal advanced earlier that infants' perceptual abilities can be deemed to be phonological. Putative parallels in lexical tone perception between infant and adult mentation are investigated. Evidence is provided that language-specific tone is perceived at around six months of age, and if this is aligned with the perception of stop contrasts we discussed in chapter 2, we have the beginnings of an acquisitional account of the primes of Government Phonology.

We set out for first base in section 4.2, which reports experiments using adult subjects. In these tests, only one tonal contrast appears to be perceivable outside of a prosodic structure, which is, as signposted in chapter 1, a completely unexpected finding, and gives us cause to add an extra hypothesis to subsequent experiments. In addition, it motivates a discussion of Yorùbá tonal typology (which is pursued in chapter 5). Following this, in §4.3, original infant perception tests are reported whose results show that the discernment of a language-specific laryngeal contrastive paradigm is present by around six months. This is followed in §4.4 by a critique of the methodology involved in the present tests. Section 4.5 is a discussion of the new findings in the light of previous work. Finally, a summary is presented in §4.6.

4.2 Lexical tones in Yorùbá and adult perception

4.2.0 The central aim of this chapter is to investigate the perception of lexical tone in infancy. A necessary prerequisite of such an investigation is to gain a knowledge of the perceptual routine(s) used by adults to map from the acoustic signal to the phonological objects which subtend lexical tone.

4.2.1 Background to the adult perception tests

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, whereas the strategy employed for distinguishing vowels is prototypical in nature.[1] 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 to use as a norm for the physiological ability of the organism to discern pitch differences within the range normally utilised in the perception of Fx.

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 noted in chapter 2 the existence of categorical perception in animals other than *homo sapiens sapiens*. We also noted that most human languages exploit this ability to categorise stop consonants, but that they have no bias toward the psychoacoustic factory settings. Speakers of English, German and Cantonese utilise the lag VOT boundary (around +30ms.), which has high acoustic

---

[1] See the discussion of the work of Kuhl and others in chapter 2.

salience and is discernable by neonates and chinchilla, while 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. Yorùbá, utilising three surface tones, has the phonologically felicitous property of possessing many words which differ only in their tonal specification. Minimal triplets exist, as exemplified in (1):

(1)

bá	'meet'	wá	'come'	kí	'greet'	rí	'see'
ba	'hide'	wa	'to be'	ki	'thick'	ri	'weeping'
bà	'perch'	wà	'us'	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 in the single parameter of 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 putatively contrasting solely on the target parameter. We will see that things do not turn out to be quite this simple, because more than one acoustic factor may be found to correlate with a single phonological object. The shape of the Fx contour, among other things, may act as a cue in addition to the manipulated value of Fx in Hz.

We may find cause for further trepidation in the malleable manifested character of tones in connected speech. In the face of this, can it reasonable to isolate tonally specified syllables for testing purposes? There are some justifications for this. The first was gained before testing began from our informant's intuitions.[2] His belief is that it is perfectly well possible to distinguish words of Yorùbá 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. The second justification is found in assertions by Siertsema (1959) and Bakare (1995) that 'short syllabic utterances ... constitute the most frequent structure in the Yorùbá language' (Bakare 1995: 54). This rather mysterious proposal by Yorùbá scholars can perhaps best be construed as an indication that the language is, relative to other languages, richly strewn with prosodic boundaries, an idea that will become crucially relevant in the following chapter. If this construal is correct, then it means that isolated syllables of Yorùbá may be comparatively invariant in citation and connected speech forms. Thirdly, the case for a necessarily relative perception of pitch-cued tones is undermined by the fact that strings of identical tones can be readily identified by a native speaker 'in the absence of a syntagmatic comparison' (Ladd 1996: 61).

However, we should be aware that context indubitably *does* infuence the character of tones. Connell & Ladd (1990) report an extensive investigation of the differing realisations of Yorùbá high, mid and low tones in running speech against a backdrop of various downtrends, which in their terms range from 'phonetic' declination to 'phonological' downstep. Contiguous tones also have predictable local effects under certain circumstances, as the data in their paper show. For example, an initial high tone will be realised with level pitch, while after a low tone, high may be signalled by a rise (op. cit.: 6). Connell & Ladd finally have to remain non-committal about the nature of these phenomena: 'we cannot hope to shed any light on the systematic relationship between these effects, much less to ascribe them to phonological rules or phonetic realisation processes' (op. cit.: 20). We will briefly consider one or two of the local effects they describe during our analysis of Yorùbá tone phonology in chapter 5, but since they state unequivocally that they do not align their findings with any particular phonological statement, it is impossible to directly correlate our points of view. Meanwhile we acknowledge that we are dealing, in these present tests, with tokens of 'citation forms',

---

[2] Extra encouragement was provided by the knowledge that Professor Oyètádé, in addition to being possessed of native-speaker intuitions, is a linguist and an author of several texts on the Yorùbá language.

and it will be interesting, in the light of the polymorphic manifestations of tone-specified words in connected speech, to see if these are sufficient to cue perception of lexical items.

4.2.2. Recording and analysis of stimuli

Recordings were made of natural utterances of the following three words of Yorùbá which differ only in tonal specification:

(2)

kí
ki
kì (as in (1))

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 the *sfs* programs[3]. These allow, among other things, audio reproduction, together with a visual display of both the amplitude signal and the spectrograph. Durations for the twenty tokens of mid and low-toned words varied minimally between about 205 and 220 ms, with the high-toned words being measurably shorter, at an average of about 160ms. 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. Stereotypical speech-pressure waveforms and spectrograms for all three tones are presented in Figures 4.1, 4.2, and 4.3.

---

[3] See §3.2 (p.67).

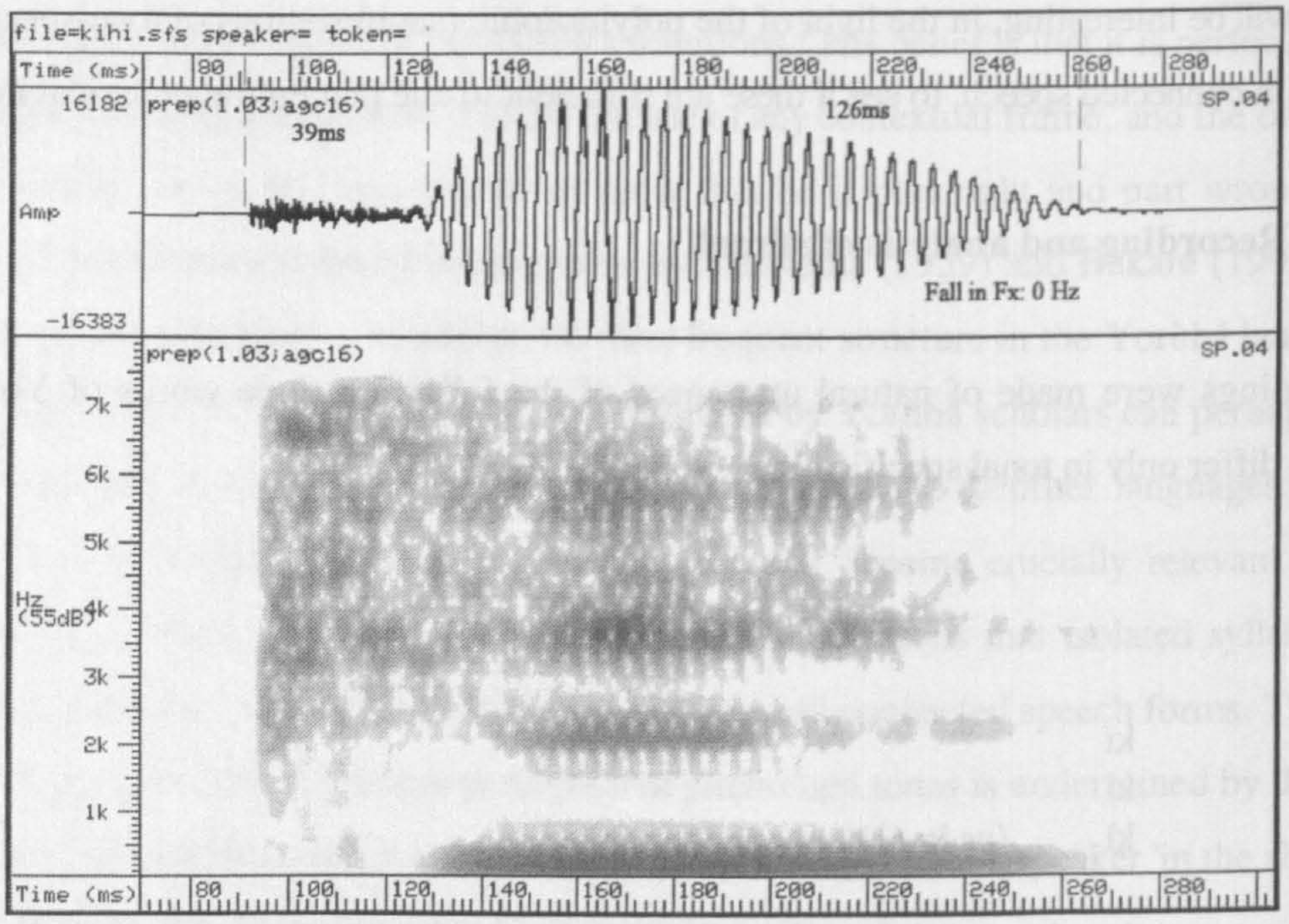


Figure 4.1 Natural token of the Yoruba word /kí/ *greet*.

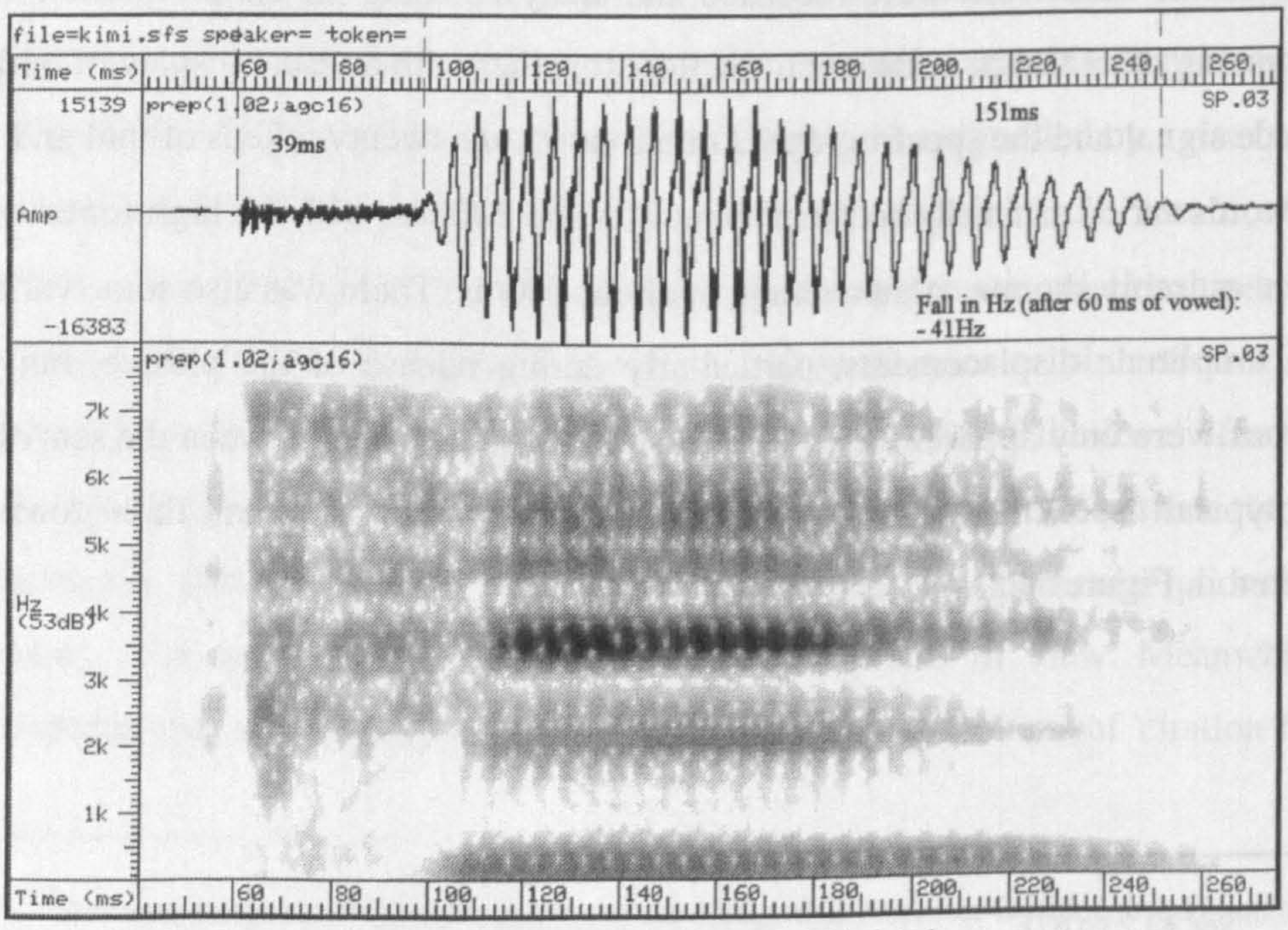


Figure 4.2 Natural token of the Yoruba word /ki/ *thick*.

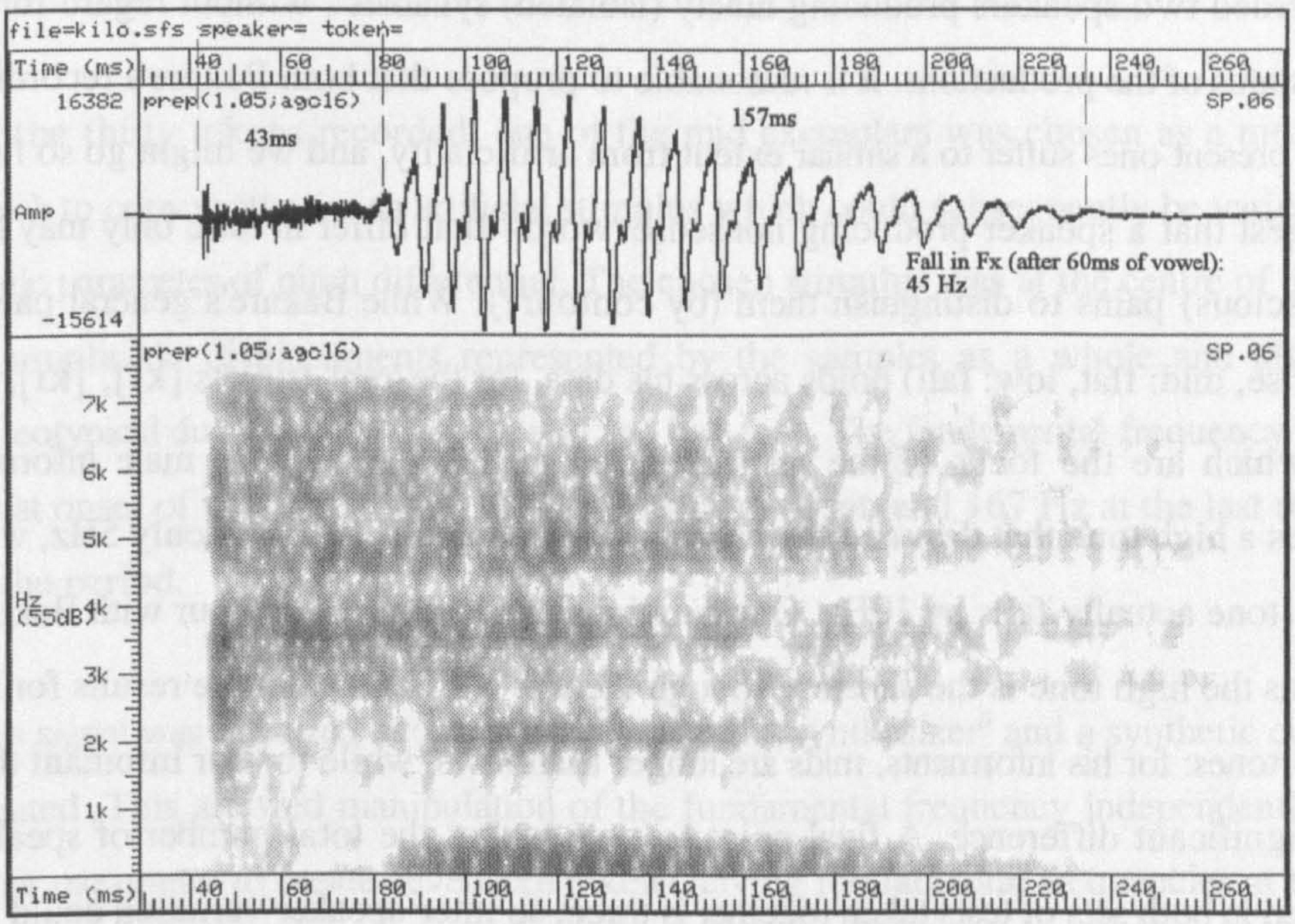


Figure 4.3 Natural token of the Yoruba word /kì/ *praise*.

The pitch variation, unlike duration or amplitude, clearly demarcated each individual word. The average values of the fundamental frequencies (in Hz) at the start of the vowel for each word are given in (3):

(3)

kí 227 ki 194 kì 149

The productions of none of the words departed more than about 3Hz from these average values. The mid- and low-toned words showed a fall in fundamental frequency of 30-50Hz toward the end of the signal after maintaining a flat pitch contour for 60-70 ms. This contour was not present in the high-toned words.

From the acoustic analysis above, it can be seen that one respect in which the pitch contour of high-toned words differs from that of mid and low-toned words is that the former lack a fall. By contrast, the tokens recorded by Bakare (1995) have falling contours only in low-toned words, high toned words are (generally) signalled by a rise, and mid-tones are more or less flat. The focus of Bakare's work is acoustic/perceptual and he recorded two speakers producing ninety (isolated) syllables[4], without regard for the lexical status of the productions. It is reasonable to propose that both Bakare's recordings and the present ones suffer to a similar extent from artificiality, and we might go so far as to suggest that a speaker producing nonsense 'words' that differ in tone only may take (unconscious) pains to distinguish them (by contour?). While Bakare's general pattern (high: rise, mid: flat, low: fall) holds across his data, for the 'real' triplets [kí], [ki], and [kì], (which are the focus of the tests reported in this section), his male informant produces a high tone that dips and then rises, with an overall change of only 5Hz, while his mid-tone actually falls by 10Hz. Our durational measurements concur with Bakare's insofar as the high tone is the shortest, though there is a difference in the results for mid and low tones: for his informants, mids are longer than lows, while for our infomant there is no significant difference. A final point to note is that the total number of speakers employed in both sets of tests taken together is three, so inter-speaker variation could still be proposed as the source of any disparities.

4.2.3 Synthesis and presentation

We return to Bakare's analyses in section 4.5, but meanwhile note that having conducted extensive perception tests to factor out Fx, intensity, duration and formant cues, he concludes that 'the best single differentiator of the tonemes was the fundamental frequency' (Bakare 1995: 32). This conclusion is relevant to a consideration of the shorter duration of the high-toned samples that we obtained. Bakare's valuable analysis shows that while duration, intensity, and third-formant values 'combine collectively to the overall differentiation among tones' (op. cit.: 46), none of them comes near Fx as a vital physical parameter for the perception of tone in Yorùbá.

Following this analysis and our own informant's intuitions[5], the decision was made to factor out both duration and pitch *contour* to see if significant results could be obtained

---

[4] There was a carrier phrase, but this consisted of a preceding [sɔ pé] ('say') followed by a pause, and there was silence after the target syllable (Bakare 1995: 34), so the syllables were as good as isolated.

[5] He has stated informally that 'tightness' is the main perceptual factor in tone discrimination.

from manipulating frequency values only.

Of the thirty tokens recorded, one of the mid exemplars was chosen as a model from which to copy synthesise an artificial stimulus which could subsequently be varied for the single parameter of pitch differential. The chosen stimulus was at the centre of the range of amplitude displacements represented by the samples as a whole and also had a stereotypical duration (212 ms) for mid and low tone. The fundamental frequency was 200 Hz at onset of voicing, 200 Hz at +160 ms from onset, and 167 Hz at the last repetition of the period.

This signal was imported into the KPE80a parallel synthesizer[6] and a synthetic copy was created. This allowed 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 Yorùbá speakers (who also speak English) and one English control. They were asked to say whether the pair represented one or two words. The Yorùbá speakers were then asked to provide English translations for the words on a second replay of the stimuli. The test(s) on the first Yorùbá speaker were repeated one month later.

4.2.4 Results of 'difference' testing

The results of the 'Same/Different' judgement tests can be summarised thus:

---

[6] See §3.2 (p.69).

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

(ii) The Yorùbá speakers recognised, without exception, any pair separated by 25 + Hz as 'different'.

(iii) The Yorùbá 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.

(iv) The results for Yorùbá speakers for the 15 and 20 Hz pairs are given in table 4.1, where column 1 gives the results for the first test on speaker 1.

15Hz differential	Speakers		*20Hz differential*	Speakers	
	1	2		1	2
170 185	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

Table 4.1: Results from two native speakers of Yorùbá testing for the perception of differences between pairs of Yorùbá 'words' separated by 15 or 20 Hz (S = 'same', D = 'different').

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

(a) The 225/205 pair was perceived as 'the same'.

(b) The pair 225/200 (25 Hz differential) was perceived as 'the same'.

4.2.5 Results of the translation test

The results of the translation test undertaken by the two Yorùbá speakers can be summarised thus:

(i) The total number of tokens of each word perceived is given in table 4.2 for all stimuli presented.

	'greet' (high tone)	'thick' (mid tone)	'praise' (low tone)	TOTAL
A	24	38	0	62
B	26	38	0	64
C	25	39	0	64

Table 4.2: Translation test: total numbers of tokens of each word perceived by A: speaker 1 on the initial test, B: speaker 1 on the following test and C: speaker 2.

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. Note that 'praise' (low tone) is not perceived at all, a point to which we return in the following discussion (§4.2.6).

(ii) For every test, each stimulus above 210 Hz was translated as 'greet' (16 in all).

(iii) For every test, each stimulus below 190 Hz was translated as 'thick' (24 in all).

(iv) For the remaining stimuli, the results are given in table 4.3.

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

Table 4.3: High- and mid-tone perception by Yorùbá speakers of stimuli between 190 and 210 Hz. A: speaker 1 on the initial test, B: speaker 1 on the following test and C: speaker 2.

(vi) On tests A and B:

(a) For every pair identified as 'same', two identical translations were provided.

(b) For every pair identified as 'different', two different translations were provided.

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

4.2.6. Discussion of results

We begin by summarising, in Figure 4.4 (over), the results of the difference testing on the Yorùbá and English speakers.

As is fully predictable from the JND experiments on pure tone previously cited, the English control perceived a different 'word' if the signals were more than 5Hz apart. All and only the signals separated by 5 Hz were perceived as the same. For the Yorùbá 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 percetual 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

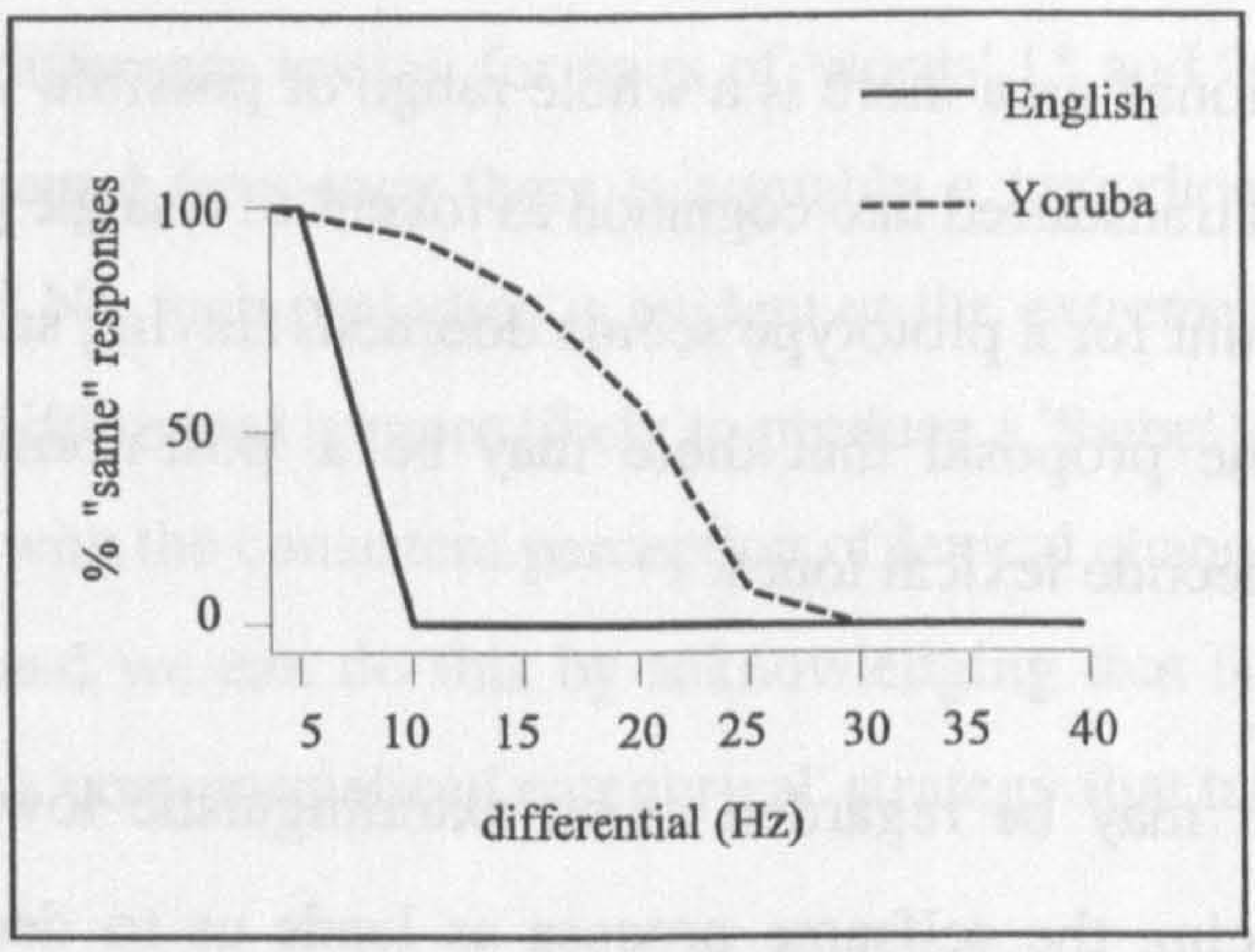


Figure 4.4: Summary of average results of Same/Different responses to pairs of Yorùbá words differing only in tonal specification by two Yorùbá native speakers and an English control.

higher in pitch than 210 Hz was always, in the translation tests, perceived as an instantiation of high tone, despite the arguable bias against this inherent in the fact that none of the stimuli maintained the flat (or even rising) pitch contour characteristic of a natural utterance. This result suggests that pitch *contour* may not be hugely salient to tone perception. Also, at less than 190 Hz a mid tone was perceived. (Low tone is different: we get to this later.) In between 190 and 210 Hz there lies a region in which confusion reigns, as can be seen from the results . 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 and prototypical respectively) 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 object [H] or to the lack of *any* phonological object (Ø). On either side of this 'transitional' area there is a whole range of possible vowel-onset pitches that can be equally well transduced into cognition as tokens of a single phonological object (or lack of it), so any hunt for a prototype seems doomed. Having said that, however, it is worth considering the proposal that there may be a 'post-normalised categorical' strategy employed to decode lexical tones.

Normalisation may be regarded 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'.[7] This is supported by the fact that normalisation is acquired extremely early in ontogeny, as attested both by linguists and psychologists. Visual size constancy is already present at birth (Slater 1992), and babies recognise their mother's voice *in utero* and so can distinguish it from any other after they are born, this being a social rather than a linguistic skill (Fifer & Moon 1989). Reference was made in chapter 2 to Kuhl's (1979) paper which demonstrates, among other things, that six-month olds identify tokens of /i/ or /a/ as members of two discrete and mutually exclusive sets, independent of pitch contour. Lastly, it is worth remembering that normalisation is also common to musical enculturation. In music, the only other structural system that we acquire from acoustic evidence, 'Rock-a-bye Baby' is 'Rock-a-bye Baby', whether it be sung in Ab or D#. So plenty of justification exists for proposing that we are 'normalising' sensory input before we utilise knowledge that is specific to any cognitive structure (such as a language), using the word 'before' in both the processing and developmental senses.

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 acoustic context. Consider the results in

---

[7] See also §2.2.2, (p.46).

Table 4.1 (p.94 above) of difference testing for pairs of 'words' 15 and 20 Hz apart. At around 200 Hz of fundamental frequency there is arguably a prejudice to perceive a change in lexical category.[8] 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 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 acoustic 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. This unexpected finding delivers up an investigable hypothesis that will be used as a new *leitmotif* for the remainder of this chapter and for all of the following one, namely:

(4) *No low tone*

> [L] is not present in Yorùbá nuclei: what has hitherto been considered as the perception of [L] is in fact the perception 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 delivers any real advantages. If true, it means we must rethink some previous phonological analyses of the languages of western Africa. But interestingly, we are presented with a reason for the asymmetry noted in §2.4 (p.60), namely that there is not a single African language that has been analysed using nuclear [L] only. In chapter 5, we look more closely at the distributions and alternations of tones in Yorùbá (and related languages) and align these facts with the *no low tone* hypothesis.

---

[8] The '200-220 S' judgement does not fit in with this proposal, but it is alone in this respect.

In the meantime, it is worth noting, albeit cautiously, that the acoustic asymmetry of the high-toned word with respect to the others in the present tests, shown clearly in Figures 4.1, 4.2 and 4.3 (pp.90-91), neatly parallels the proposal in (4) about Yorùbá tonal phonology. At least for our informant, linguistic knowledge requires high tone, in isolation, to be realised with an acoustic shape quite different from either of the other surface tones.

The experiment described in this section could certainly be extended. 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 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]. It may also be important to carry out some translation tests on single words (perhaps separated by music), and it could be interesting to have subjects rate the quality of various pitches as instantiations of tone(s). None of these interesting potential extensions of the adult tests, however, debar our pursuing parallels between adult and infant perception.

So in the present study, encouraged by the accord between the above findings and the typological asymmetry presented by the phonological analysis of African tones in general, we will add *no low tone* to our other testable proposals (§2.4, p.57) for infant perceptual experiments. It is to these that we now turn our attention.

4.3 Infant perception experiments

§4.3.0 We have discussed (§2.2) previous tests that demonstrate the sensitivity of six-to-eight month old infants to source contrasts in obstruents which are specific to the language they are acquiring. If these perceptions map to melodic phonology, they display an attunement of perception to the particular distribution of the elements [H] and [L] in the target language. It has been demonstrated in §4.2 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 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. This section describes a series of infant perception tests carried out with this aim in mind.

4.3.1 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. The paradigm comprises a *Training Stage* and a *Testing Stage.*

The experimental site is arranged as shown in Figure 4.5.

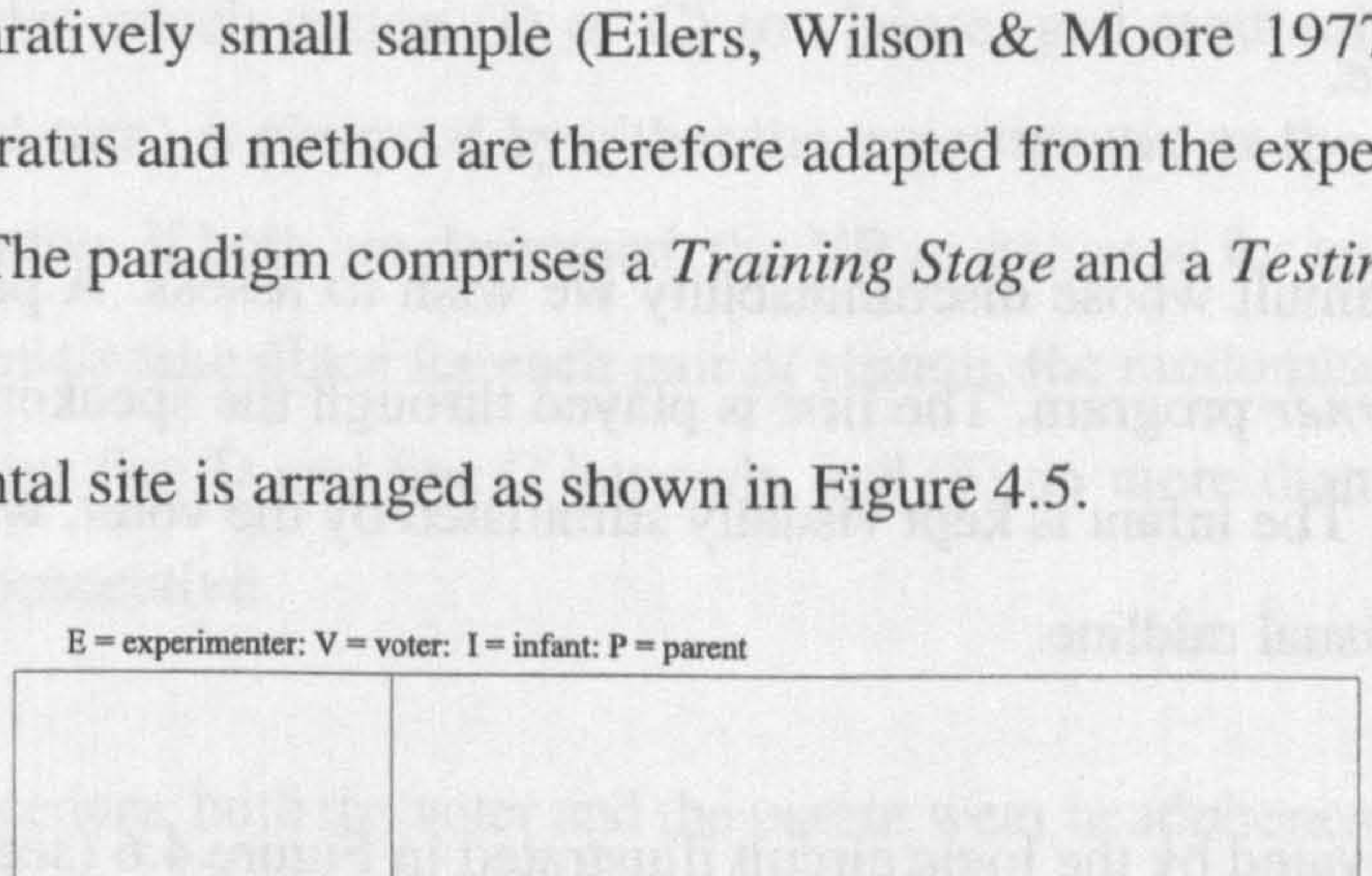


Figure 4.5: Experimental suite for infant perception testing.

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[9] 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.

During the *Training Stage*, a pair of two easily discriminable stimuli are chosen. These are loaded into the *Turner* program[10] in the PC, which replays the first at a regular interval of one or more seconds.[11] 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 has been achieved, 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 in Figure 4.6 (see Eilers, Wilson & Moore 1977: 769).

---

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

[10] Developed by Mark Huckvale at UCL.

[11] 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).

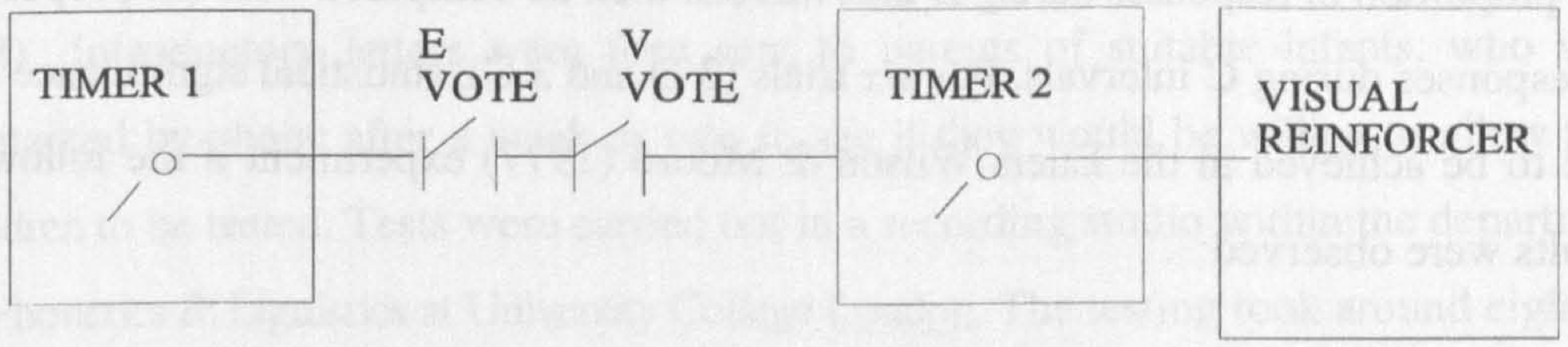


Figure 4.6: Logic circuit for activation of reinforcer during the *Testing Stage* of the experiment.

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.[12]

---

[12] Sample *Turner* logs are presented in appendix B.

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 (1980a) 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.

4.3.2 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.[13]

4.3.3 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 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 §4.3.4 below. These *Training Stage* pairs 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

---

[13] To comply with legal requirements, ethical approval for testing was obtained from the *Joint UCL/UCLH Committees on the Ethics of Human Research* (ref: 96/36).

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.

The synthetic tokens of the 'words' [kí], [ki] and [kì], previously used for adult testing (§4.2.3 above), 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 225 Hz, and vary by 5Hz. New stimuli at 230 Hz and 240 Hz were manufactured for use during the infant tests. 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.[14] It was therefore 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 §4.3.4 below), used all three differentials, and the main study 20 Hz only.[15] Reliable discrimination of any pair less than 10 Hz apart by any child was neither expected nor predicted off our adult perception tests (§4.2.1 and §4.2.4 above): 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.

4.3.4 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

---

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

[15] With one or two minor exceptions: see §4.3.5, (p.111).

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[16] (§4.2.6, p.97 above), 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.[17]

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 4.4: 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.

---

[16] This is of course, as we have said, a post-normalised value.

[17] 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.

In neither case did the English subject respond in a systematic way to any of the pitch-differentiated stimuli.Yorùbá infant 2 appears similarly uninterested, apart from 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 physical disruption by aggrieved concert-goers, 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.[18] 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 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[19], 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

---

[18] 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.

[19] This idea is insupportable in the light of the assembled research programs cited in chapter 2.

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 (§4.2.6, p.99) 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.

(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.

4.3.5 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.[20] 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 (5):

(5)

(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 §4.3.4 p.106 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

---

[20] The role of non-Yorùbá language input is discussed in §4.3.2 (p.104) above.

Testing Stage: if not, an attempt was made to train them on 'bà' and 'kí' before proceeding. *Training Stage* results are presented in Table 4.5. *Testing Stage* stimuli were paired tokens of the syllable 'ki' which maintained an Fx differential as set out below in Table 4.6, 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 4.5: 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).

	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 4.6: 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%).*

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 4.6 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.[21]

For these reasons, the results are hardly as conclusive as we could wish, but three interesting patterns do emerge from Table 4.6., 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.[22] These regularities merit an explanation. We will consider them in the light of the questions in (5) above.

Question 5(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 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 (§4.3.3, p.106).

---

[21] See appendix C for the raw numbers.

[22] 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.

A precursor to the answer to the question posed in 5(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) contradict this proposal. 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 5(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[23], 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 adult Yorùbá. This is at least consistent with the assertion in (4) (§4.2.6, p.99), which proposed 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.

---

[23] See §2.2.2, (p.46) & §4.2.6, (p.98).

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).[24]

An alternative analysis of these findings is presented in Figure 4.7, (over) where the number of successful responses within each of the categories *190/210 Hz, 210/230 Hz* and *all other stimuli*[25] is expressed as a percentage of the total number of tests for that category.The story so far inspires both further empirical investigation and the search for support for these findings from independent sources. The latter of these aims is the focus of the next chapter, which is an analysis of Yorùbá tonal typology. In §4.5 and §4.6, we look at the results reported here in the light of extant work. As for further testing, 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-à-vis* phonological perception could be repeated.

---

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

Yorùbá group:		χ^2 Values
	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

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

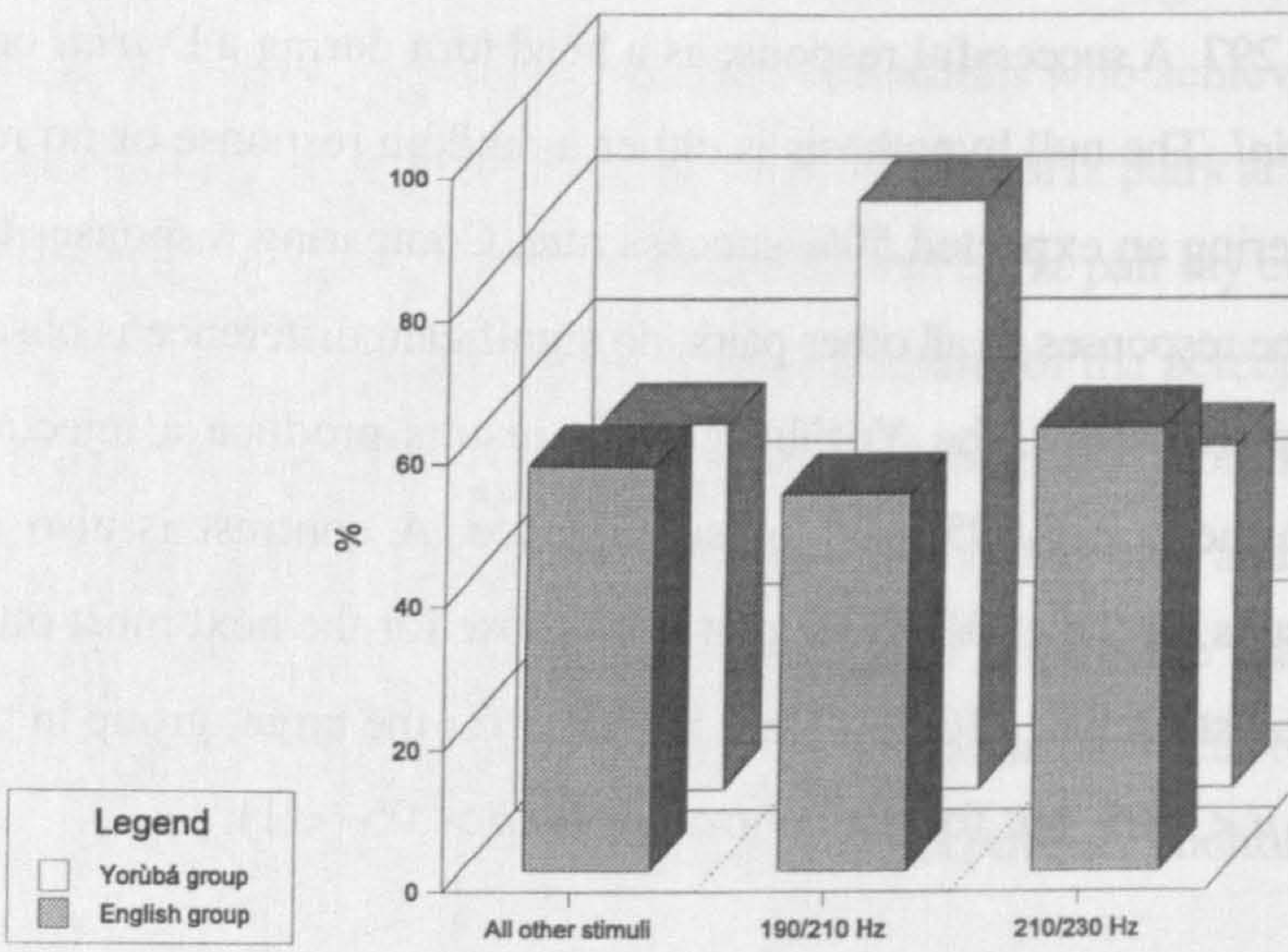


Figure 4.7: % successful responses for each group on (a) all stimuli except the 190/210 Hz pair, (b) the 190/210 Hz pair, and (c) the 210/230 Hz pair.

For example, there could be detectable acoustic properties present in the *natural* low tone utterances which cue the perception of [L].[26] Bakare's previously-cited work (1995: see §4.2.2, p.91) does not, however, support this assertion, and finds that Fx is the overwhelmingly significant cue to the perception of 'tonemes'.

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. But first it is responsible to note sources of potential experimental error in the present tests, with the particular aim of improving the reliability of future results from infant perception experiments in general.

---

[26] John Coleman (pc) pointed this out. It would be possible to produce synthetic tokens of words having 'breathiness' which could be utilised for the future testing of this proposal.

4.4 Critique of the VRISD paradigm

4.4.0 In §2.2.3, the advantages of the Visually-Reinforced Infant Speech Discrimination paradigm over other techniques (specifically HAS) were discussed. Among these was the ability to get significant results from a very small sample, a quality that has been exploited in the present work. Because of the attractivness of the Visual Reinforcer (to infants), VRISD also circumvents the problem of subjects discriminating sounds but being uninterested in the discrimination, a problem which preferential listening experimental routines must take into account (Plunkett & Bryant 1998: 4). However, during the course of carrying out the tests reported in §4.3, it has been possible to detect certain shortcomings implicit in the method, and to compare this particular VRISD site both with another visual-reinforcement setup and with a non-behavioural test (ERP[27]). A comparison of all three is offered here with the aim of assisting the design of future enterprises of this type.

4.4.1 Shortcomings of the present VRISD paradigm

This method has a convincing ability to withstand human 'interference' in the recording of head turns. The headphones worn by the assistant and the infant's carer, together with the randomiser built in to the program, ensure that none of the adults present know if the stimulus changes or not. An extra safeguard exists in the double vote necessary to record a positive head turn. A small problem occurs, however, when the tone is played into the *voter*'s headphones to alert her to the onset of a trial. This is bound to change her attitude to the infant, who, though focussing on the visual midline, is well aware that the Visual Reinforcer (VR: toy penguins, in the present experiment) could be activated at any moment. The infant's potential recognition of this change of attitude, together with his or her anticipation of the VR, could certainly bring about a head-turn.

Another problem lies in the impossibility of controlling for the infant just 'having a look' to see if the penguins will come back. This certainly happened time and again in all the

---

[27] Event-Related Potential measurement. See §4.4.2.

tests, quite irrespective of anything that was being played through the loudspeaker. As long as this did not occur during a trial, the procedure ensured that the computer would not log it as a positive. However, it is perfectly possible that it could coincide with a trial, in which case the two votes would be received and the VR activated.

Both of these problems are linked with the hyperefficiency of the VR. Once the children had seen the penguins once, it was perfectly obvious from the way they behaved that they knew exactly where the VR was, and that they would like to see it again. The *voter*'s hardest task was to get the subject to forget about the VR long enough to face forward again, and stop staring at the blank space where the penguins had been. Other researchers (Rivera-Gaxiola, pc) have found that some babies are frightened of the VR: this was not a problem encountered during the tests reported here.

Even the most calm and attentive subject had, on a good day, the ability to concentrate on the task in hand only for a limited amount of time. There is thus a danger that, before any of the adult observers becomes aware of the boredom or fatigue of the infant, he will have become wholely uninterested in the speech stimuli and will begin to register negatives to changes he is perfectly well able to discern. This is related to a further problem. Some of our findings show a child scoring correctly on the first few trials and then failing on the last. Here we may have evidence of positive discrimination blurred by external mental events. There are so many uncontrollable factors in considering the child's general attention, including our inability to characterise exactly the relation between adult consciousness and its semi-developed state in infancy, that getting at computational processes from this (behavioural) direction is fraught with uncertainty.

Finally, the role of the *voter* in holding the infant's attention is not necessarily as neutral as could be desired: being human, the *voter* understandably forms some kind of relationship with the infant which varies from moment to moment, and so arguably influences the subject's responses.

Future adaptations could be made to the VRISD paradigm to overcome some of the above objections. First of all, it is possible to videotape the child's responses, and have the tapes viewed and the responses coded by independent observers, a process which is now in use in a laboratory in France (Rivera-Gaxiola, pc). It is also possible to substitute a non-human device for the *voter* in order to hold the subject's attention at the visual midline, and to remove the parent/carer from the suite. Current tests at the University of Barcelona are using a VDU displaying changing coloured swirls in front of the infant, who sits in a chair, rather than on the carer's lap (Bosch, pc). As long as the baby is calm at the onset of these tests, the lack of interaction with other humans in this environment removes many sources of potential distraction or influence. A similar VDU device can be used in non-behavioural testing, a paradigm which also removes the problems associated with the nature of the VR. It is to non-behavioural tests that we now turn our attention.

4.4.2 Non-behavioural infant speech perception testing

Developments in non-invasive methods of measuring brain activity directly by electrophysiological methods could potentially by-pass some of the problems encountered in the present tests. Event-related Potential (ERP) testing as described, for instance, in Woods and Elmasian (1986) demonstrates that voltage at particular sites in the brain can be depleted by as much as eighty percent on the second or third repetition of a stimulus, and that recovery of potential occurs on the introduction of a new stimulus. Measuring infant responses to a passive linguistic task in this way, speech stimuli being played while the child's attention is being held elsewhere, may give us a clearer view of unconscious processes than has been possible with the present unmoderated behavioural procedures.

One crucial question to face in assessing ERP results is to decide exactly what kind of perception triggers the voltage change. Dehaene-Lambertz & Dehaene (1994), using ERP tests on infants to demonstrate that at two to three months a child will respond to a change from [ba] to [ga] within 400ms, note that their reported ERPs exhibit three different waveforms. Rivera-Gaxiola (pc) has gone as far as proposing that 'it might be possible that each one reflects a stage of processing of the signal, namely acoustic, phonetic and phonemic'.[28] While it is obvious by now that, in the terms adopted in this discussion, processing stages cannnot be subdivided in this particular way, the crux of Rivera-Gaxiola's suggestion opens up possibilities for a direct approach to some of the *leitmotifs* considered here. If there is *any* kernel of truth in her acknowledgedly speculative idea, then babies could be shown to be 'doing' phonology, a task impossible without independent phonological primes (and, incidentally, definitionally impossible under the banner 'phonemic'). Such a demonstration, while worth striving for, seems to be some way off. Rivera-Gaxiola's own ERP tests on adults (Rivera-Gaxiola 1998) illustrate the problems. She first demonstrated behaviourally that a voiced stop continuum running from bilabial through coronal to retroflex is perceived differently by speakers of different languages. Native English speakers discriminate [b] from (alveolar) [d] (categorically), while native Hindi speakers discriminate three items, namely [b], [d̪] and [ɖ]. She then followed these tests up with ERP studies which broadly showed that while 'acoustic sensitivities, categorical perception and category goodness contributed to the waveforms obtained,... covert discrimination capacities were still present for all the conditions in these studies' (Rivera-Gaxiola 1998: 2). This ubiquitous sensitivity to acoustic stimuli leads Rivera-Gaxiola to hold, among her conclusions, the notion that 'the brain perceives more than the mind can'. We might wish to take issue with the way this is put, as perception is a property of the mind/brain (in the Chomskyan sense) rather than of the physical organ. Moreover, perception, exceptionally, requires no external input (hallucinations, dreams...), and as Rivera-Gaxiola points out (op. cit.: 174), there may well be a need to take into consideration some kind of perceptual retrieval system which is used to match to a given acoustic input. What will be needed to make ERPs really useful in this kind of test is some more sophisticated way of analysing and comparing the character of individual waveforms.

In the meantime, it would be interesting to bolt an ERP setup onto the VRISD paradigm, and see if a recognisable style of voltage increment was associated with a headturn. Here, another real problem remains, prosaic though it may be. Extant ERP sensors are

---

[28] This suggestion was circulated to several interested parties by Maria Rivera-Gaxiola while she was working on her PhD at the MRC Cognitive Development Unit, London.

cumbersome, and *look* invasive: none of the parents approached during the present VRISD tests would even consider putting their babies into such a contraption. If the ergonomic design of the testing equipment can be improved, then the advantages of VRISD mentioned in §2.2.3 and §4.4.0 may be harnessed to the objectivity possible in non-behavioural testing.

For the present, early acquisitional research can proceed primarily by aligning such findings as are obtainable empirically with the most convincing theories of mind: this enterprise conforms to the Jakobsonian ideal alluded to in §1.1 and §2.1 above.

4.5 The present tests and previous experimental work

The infant testing program described in §4.3 is the only extant work on the perception of lexical tone in the first year of life.[29] In its inception, it was substantially motivated with the objective of aligning lexical-tone perception in infancy with known findings on stop perception[30] insofar as both modes of perception could be seen to have developed in a chronologically parallel fashion. This, it was argued, 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 have actually answered these questions simultaneously. Training subjects on linguistically viable stimuli, we have argued, must awaken the expectation of linguistic discriminations 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

---

[29] Demuth's (1993) Sesotho study, briefly alluded to in §2.4 (p.60), used production data from a Basotho boy over the twelve month period from 2:0:0 to 3:0:0.

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

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.* At around six months of age, the language-specific sensitivity to the perceptual cues for [H] and [L] discussed in chapter 2 is paralleled by evidence of a developmental change in the minds of acquirers of a lexical tone language. But it has now become even more desirable, if we are to further shore up this claim, to show that Yorùbá tonology has one, rather than two, underlying lexical tones: chapter 5 confronts this idea.

Turning to previous work on adult perception of lexical tone, the most substantial report on Yorùbá is found in Bakare (1995 and references therein).[31] We have already discussed the degree of concord (and discord) between his acoustic measurements and the present ones, and his work was used here to partially justify both the employment of isolated syllables for testing, and the sole use of Fx as a variant parameter. We now turn to his perception-testing results and attempt to align them with our own.

Bakare used three different control groups. In addition to Yorùbá and non-tone-language groups, which correspond with our listeners, he also used a group of Chinese[32] native-speakers and a group of trained musicians. He partially summarises his results: 'when results were ... pooled across all linguistic variables ... a perceptual hierarchy was shown in the recognition of the tones, thus: Yorùbá > Chinese > Music > Naive. Linguistic experience improved accuracy more than non-linguistic experience' (Bakare 1995: 37). Bakare's subjects were required to identify or discriminate natural utterances of one tone from another, which is of course a different kettle of fish to the task set to the subjects in the present adult tests, who had to try and identify or discriminate syllables that varied in pitch only, and sometimes by as little as 5Hz. Bakare's subjects therefore had a far easier perceptual task. Many acoustic cues (including contour) must be present in the stimuli he uses, and the seminal cue, Fx, varies by as much as 100Hz between high and mid tone (at

---

[31] See also §4.2.2 (p.91) & §4.3.5 (p.111).

[32] Bakare doesn't say which 'Chinese', but whichever it was should not affect the outcome.

end of [k͡pu]: female speaker) and 110Hz between mid and low (at end of [su]: female speaker).[33] It is unsurprising that accurate discriminations were easier to come by in Bakare's work, 'without any help from linguistic context' (op. cit.: 46). The perceptual hierarchy referred to above reflects the value of non-linguistic pitch training for the musicians, as Bakare says, but does it not also do the same for the Chinese group? They could not possibly have been, like the Yorùbá subjects, 'assumed to be in a completely linguistic mode' (op. cit.: 47), and were arguably generalising from attuned cognitive skills to attend to the task they were set. As the results show (op. cit.: 36) the Yorùbá group possesses recognition capabilities that are statistically better than all other groups for both discrimination and identification. This is the only inter-group statistically exceptionless result: in particular, differences between the Chinese and Music groups are not significant, as Bakare points out, for the discrimination experiment.

A real exploration of how far the Chinese group are capable of phonologically accessing Yorùbá tones should use a serious comparison of the tonology of the two languages. To approach this in any detail is beyond the scope of the present discussion, but however the six surface tones of Cantonese or the four of Mandarin are analysed, and whatever perceptual routines may be discovered to access the objects which result from this analysis, any acoustic/articulatory overlap between Yorùbá and Chinese cannot be assumed to be complete, or other than arbitrary, without such a comparison. A study of this kind would need, among other things, to take account of the fact that these particular Asian languages differ from their west African counterparts in that they possess inherently contoured tones.

The conclusion we would like to draw from all this is that it is *phonological* knowledge that makes the Yorùbá group uniformly superior at these tasks. Two other findings reported by Bakare may bear on this question.

The first is the 'superior left ear' effect detected in the Naive group of listeners in Bakare's tests. For normal right-handed subjects, several aspects of language processing are right

---

[33] Figures from Bakare (1995: 41).

ear (left hemisphere) dominant (Kimura 1961), while music is generally left ear (right hemisphere) dominant[34] (Kimura 1964). Bakare plausibly concludes that the Naive group are processing tones non-linguistically, in contrast with the Yorùbá group, who exhibit no ear asymmetry. The problem comes when assessing the fact that, like the Yorùbás, neither the Chinese nor the Music group display asymmetry. From Bakare's point of view, that the Chinese group does not is unsurprising, but he finds the musicians' symmetry 'paradoxical', and holds the opinion that future testing may 'induce a left ear advantage for both Music and Naive groups and a right ear advantage for Yorùbá and Chinese listeners' (op. cit: 48). There is a psychological objection to this. Pitch perception may be mapped to music, language or neither. The possibility has already been raised that the superiority of the Chinese *and* the Music groups' results to those of the Naive group is due to a general cognitive attunement of pitch perception, even though pitch perception in each case has been enlisted in the service of a different structural system. This attunement may result in a loss of measurable left ear bias for all concerned. The possibility of further testing in this area is an interesting one, but the outcome is not a foregone conclusion. It is impossible to be confident about the relevance of experimental results to different 'levels' of perception or 'modules' of cognition without shoring them up with independent research. Precisely this is attempted in chapter 5, where support for the results presented in this chapter is sought in typology.

One final aspect of Bakare's findings needs to be aligned with the adult perception test results presented in §4.2. Bakare reports the existence of a switch in the perceptual hierarchy for tone exhibited by the Yorùbá speakers. The analysis of errors performed by Bakare revealed that all three control groups ranked the relative ease of tone perception as high > mid > low, 'the natural psychological dimensions' (op. cit.: 47).[35] By contrast, the Yorùbá group ordered the perceptual 'ease' hierarchy high > low > mid. The simplest possible explanation for this asymmetry is the harnessing of phonological knowledge to other cognitive abilities, and the finding supports our assertion that perception by the

---

[34] Like other percepts which, in Jackendoffian terms, may not reach the 'conceptual' module.

[35] A similarly non-linguistic explanation must be given for the fact that all control groups found the tokens produced by the female more comprehensible than those produced by the male.

Chinese group cannot be considered to be phonological. If we accept that the switch is accounted for by native-speaker linguistic knowledge, we run up against the straightforward contradiction that Bakare's Yorùbá subjects did perceive the low/mid distinction while the subjects reported in §4.2 above did not. How may we resolve this paradox?

It is possible, as acknowledged in §4.2.6 (p.100), that our own stimuli lack the non-frequency-related acoustic cues needed for the discrimination of low and mid tones. But at least a part of the resolution of the paradox must lie in the relative ease of general discrimination of Bakare's stimuli, and more importantly, his experiments may not focus the minds of his (Yorùbá) subjects closely on lexical items. Since the tests reported in the present work, in particular the translation test, demand attention to the stimuli *as words*, an assurance is obtained that our subjects are in 'linguistic mode'. Support for the suggestion that Bakare's results reveal a different 'mix' of perceptual skills from those cited in §4.2 comes from an encouraging parallel found in his remarks concerning his discrimination experiment. *Only* 'the high tone was better perceived than either the mid or the low tones' and 'the high tone was significantly different from the mid and the low tones' (op. cit.: 37). Though the perceptual hierarchy was 'switched' in the case of the Yorùbás, statistical significance was only obtained for the difference in the number of errors relating to high *versus* non-high tones, not for mid *versus* low. Bakare's work concurs with the results given in §4.2 insofar as the discrimination of high from non-high tones is a more robust ability for native speakers than discriminating among non-high tones. The stimuli he uses contain sufficient information to distinguish high tones from others (using a mix of general cognition and phonology), but there is evidently not enough information present to discriminate mid from low at a comparable level of reliability.

The switch in the perceptual hierarchy may also be viewed as the result of employing language-specific phonological knowledge in a rather different way that is also consistent with the analysis of Yorùbá to be presented in the following chapter. Bakare did use a minimal carrier phrase - [sɔ pé __] ('say __) - so despite the artificial isolation of his stimuli relative to running speech, there may just be enough prosodic information present to have consequences for some of the perceptual results.

4.6 Summary

This chapter has made the following assertions, based both on the experimental results given in §4.2 and §4.3, and on earlier work:

(i) The perceptual routine for lexical tone is categorical (but of course post-normalised).

(ii) Infants tune into the lexical tones of their native language in the first year of life.

(iii) The identity of the phonological objects underlying stop contrasts and lexical tones is supported from a chronological point of view in acquisition.

(iv) Given that both the adult and infant subjects were using phonological perception, nuclear [H] and [L] are not congeneric objects in the Yorùbá language.

The last of these claims is, from the point of view of our test results, the shakiest, because of the possibility that the perception of [L] may not be cued as exclusively by pitch as is the perception of [H]. In this context, some interesting recent research (Hayward, Watkins & Oyètádé 1998) displays results that suggest that Yorùbá low tone is associated with a different phonation type than that associated with other tones. 'Other things being equal, breathier phonation occurs on the H and M tones, whereas creakier phonation occurs on the L tone'[36]... the H and M tones show a smaller degree of [spectral] tilt' (1998: 9). If voice quality serves as a perceptual cue for the discernment of low tone from others, this

---

[36] The correlation of 'breathy' with (at least) H and 'creaky' with L might profitably be linked with tonogenesis: (see §2.4, p.59).

would help to explain some of the results reported in this chapter, since no adjustments were made for this acoustic parameter.

The phonology of Yorùbá tone grouping (and of African tone hierarchies in general) does not, however, reveal any 'natural class' comprising H and M as opposed to L. In contrast, as we will discuss in the following chapter, there are circumstances where M and L may alternate, H displays a general 'robustness', and a H>>L>>M tendency is manifest under phonological pressure. While there is no obligation on a psychoacoustic system to parallel phonology, we may expect this generally to be the case in the light of our remarks about the nature of these systems in §1.3.2 and §1.3.4. It is therefore not unfeasible that the different phonetic characteristics of L may reflect a different phonological role altogether. A phonological explanation for the results of the perception tests is certainly worth investigating.

PAGE
MISSING
IN
ORIGINAL

Chapter 5

Yorùbá and Nuclear Laryngeal Elements

5.1 Introduction

As theoretical economy is generally a desirable goal, one conceivable aim within the element- theoretic account would be to allow all elements to freely associate within prosodic structure. This is not, however, an achievable ideal. Firstly, for a given language not all elements will be utilised in all prosodic positions, and secondly certain combinations may be widely disfavoured. The former scenario, wherein an element does not show up at all in a position, is displayed by the bifurcation of European languages[1] into two sets, those which utilise aspirated stops and no fully-voiced stops, and those which use fully-voiced stops and no aspirated stops. If there is no aspiration used, then in the model we have outlined, [H] never inheres in an onset. If no full-voicing, then [L] is debarred from this position. The second of the asymmetries which show that elements do not freely combine is exemplified by the lack of front rounded vowels in English (and in a majority of languages), which can be formalised as an instance of tier-fusion whereby nuclear [I] and [U] both occupy the 'colour' tier and are therefore not at liberty to link to the same skeletal point.

The notion of acquisitional parameters has the neatest explanatory adequacy in dealing with these departures from free elemental association. The very differences between the melodic inventories of languages may be presumed to be the result of parameter setting. Furthermore, typological majority decisions such as 'no front rounded vowels' can be presented as an OFF setting for the relevant parameter.

To delve into any particular language and bring to conscious scrutiny its melodic inventory

---

[1] 'Languages' and 'dialects' are synonymous here (no less than everywhere else).

is an enterprise which requires us to look further than the physical or psychophysical correlates. The present understanding of the phonological attributes and *modi operandi* of the source elements is derived from typological and historical studies, as well as phonetic measurement and psychoacoustic factors.

The results of the experiments reported in the previous chapter indicate that low tone (pitch) in Yorùbá may not map to lexical [L], or indeed to anything at all outside of a prosodic structure.[2] This runs counter to a reasonable first guess at an elemental account of this three-level-tone language which would logically posit [H] and [L] as the melodic objects which carry the distinction, with the third tone being phonologically inert (Ø). There are, though, 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.

Yorùbá is a Benue-Congoid language, spoken in western Nigeria and in Benin. It has a number of 'dialects', but the most widely discussed and documented version is 'Standard Yorùbá'. This is based on the Òyó dialect, a vernacular from north of Ibadan in south west Nigeria. The status of 'Standard Yorùbá' is roughly comparable to that of English 'RP' or Greek 'katharevoussa' in that it is taught at school, no longer closely associated with a geographical area, and has been defined as 'that spoken by newsreaders on the radio' (Siertsema 1959: 42). Though we largely confine our attention to data from this language (and we henceforth refer to it simply as 'Yorùbá'), we should bear in mind that the evidence of asymmetries in tonal representations that we will cite in no way marks 'Yorùbá' out as unique. The common pattern, indeed the only pattern, indicated by the phonological behaviour of nuclear tones in west African languages, demonstrates that a hierarchy exists such that:

high >>> low >>> mid,

---

[2] This finding rests easy with the fact that, to date, no African language has been analysed as 'nuclear-L only'.

where we can, for the moment, informally read ' >>>' as 'is more robust than'. This hierarchy will be exemplified several times in the forthcoming sections, and ultimately we will try to align it with the *No Low Tone* (NLT) hypothesis proposed in the previous chapter, i.e. that the pitch perception previously presumed to cue [L] in Yorùbá is in fact associated with a prosodic constituent. This entails a discussion of quantitative maxima and minima in §5.3, which reveals Yorùbá to be an exotic language in this respect. Within its unusual prosodic architecture, licensing constraints on certain elements are to be discerned (§5.4). NLT gives us the opportunity to frame a general constraint against laryngeal elements in dependent nuclei. We are also led to consider the existence of a suprasyllabic, subword prosodic constituent for the language. In §5.5 some previous analyses of related languages are reviewed in the light of the proposals in this chapter, and NLT is then set in a general context of 'audible prosodic boundary' (§5.6). A summary follows in §5.7.

Before getting down to the particulars of Yorùbá tonology, though, the subsequent section cites independent sources to substantiate that idea that *presumptions* of [H]/[L] phonetic symmetry are generally unsafe, and in particular that [L] does not need to be componential in the phonology mapped to low pitch.

5.2 [L] and phonetic signatures

5.2.0 One of the *leitmotifs* of this thesis is the idea that the same element may be associated with disparate phonetic signatures, and that this disparity is a result of its prosodic assignation. While the correlates of [H] and [L] are commonly laryngeal, they show a diversity (as far as we have yet pursued this matter) which ranges from pitch changes to aspiration. Our wariness about using phonetic generalisations as an exclusive tool in assigning phonological identity is concordant with this proposal. Some recent research has had the effect of further loosening the theoretical ties between phonological symmetries and phonetic predictability. This section will examine the notions that (i) low pitch may have linguistic significance without elemental content (§5.2.1), and (ii) [L] may map to nasality as well as 'full-voicing' (§5.2.2).

5.2.1 Boundary tones

Pierrehumbert (1980)[3] delimits a system of intonational taxonomy which has subsequently been highly influential on research in this area of phonology. The notation she uses represents the intonational contour as a series of pitch accents, which comprise either a single H or L tone or some duple combination of these, together with edge tones,[4] which are either H or L. A system using such primes emphasises that there is no fundamental difference in phonological structure betweeen lexical-tone and intonation (or pitch-accent) languages. The difference is rather 'how the tonal specifications come to be where they are' (Ladd 1996:149): they may be either a part of the phonological address of a morpheme, or they may be postlexically assigned. Unlike the traditional British intonational model[5], which uses primal units like the (intonational) Nucleus, the Pierrehumbert system is fairly transparently compatible with the element-theoretic account. However, in a genuinely privative framework, it should be noted that (at least) Pierrehumbert's boundary tones may already be overspecified in some circumstances, because unless evidence of phonological activity (such as spreading) of two elements is available, a two-way *contrast* only logically requires one prime.

Cabrera-Abreu (1996) takes this basic arithmetical statement and expands it into a plausible account of English intonation using only one tone. This tone she notates as [T], but it is explicitly identified with the H of the Pierrehumbert system. L is made defunct for a variety of reasons ranging from its 'default' role as a boundary tone to its failure to participate (in lexical tone languages) in phonological processes. The single remaining tone interacts with other theoretical constructs to derive the variety of English intonational contours. These constructs are prosodic: [T] is assigned at projected nuclear levels and

---

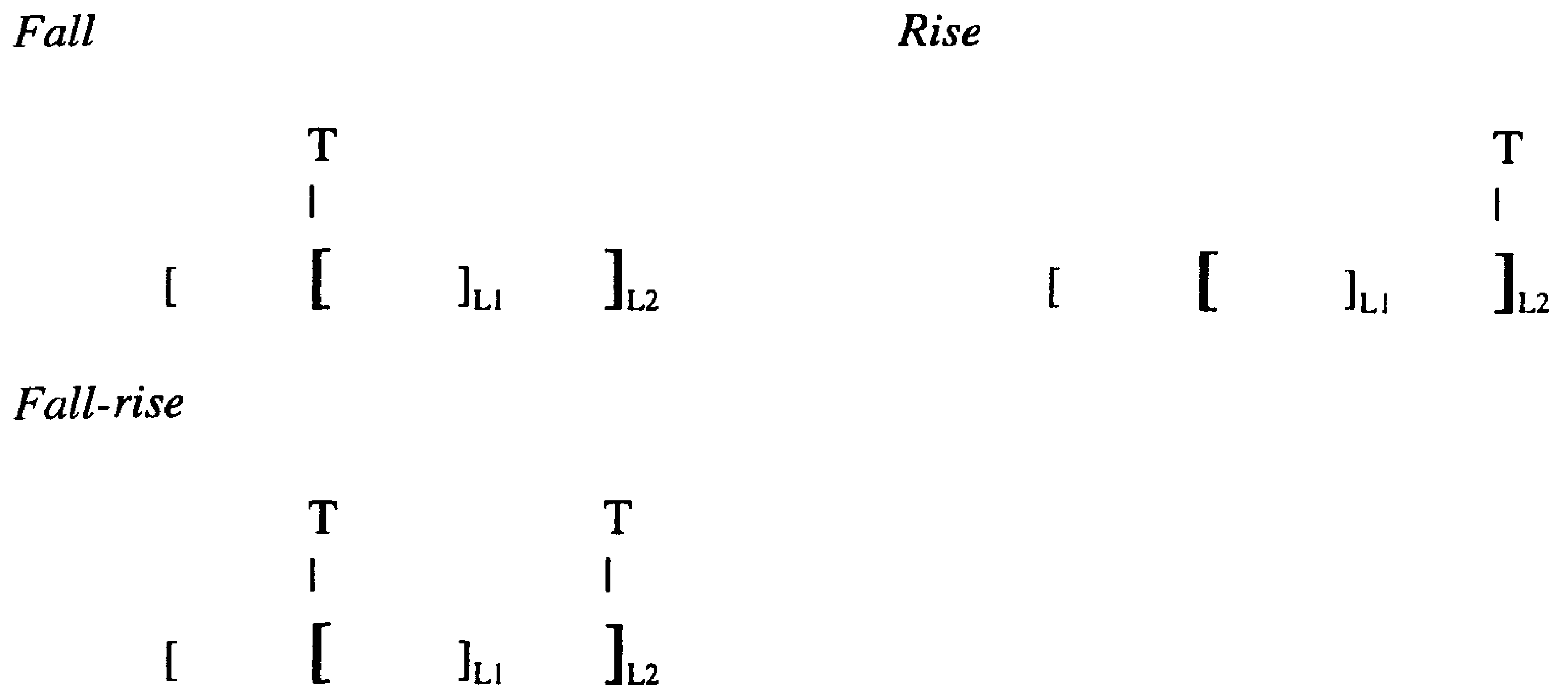
[3] And also as adapted in Beckman & Pierrehumbert (1986) and Pierrehumbert & Beckman (1988).

[4] Some versions of the Pierrehumbert model use 'phrase tone' and 'boundary tone' at the edge of different sized phrases: these distinctions are not of concern in the present context.

[5] As formulated in O'Connor & Arnold (1973).

must be assigned in conformity with the directionality of licensing (1996: 87). We therefore end up with formal representations of pitch patterns where low pitch is not the result of elemental activity, but is instead a pattern associated with the presence of a prosodic boundary. The three English 'Nuclear Tones' often regarded as 'basic' are illustrated in (1) (from Cabrera-Abreu 1996: 84).

(1)



The subscript 'L's here indicate two different levels for 'boundary licensing'. This licensing displays the same directionality observable elsewhere in the grammar: L to R at level 1 (cf. foot licensing) and R to L at level 2 (cf. word licensing). The larger, emboldened square bracket therefore represents the licensor position at each level, and in conformity with licensing inheritance (Harris 1997) the reduced potential of the dependent positions debars them from a-licensing [T].[6] [T] is thus associated exclusively with prosodic head positions: as we will see, this is exactly what happens to [H] in Yorùbá (§5.4.1, pp.147ff.). Note in particular that Cabrera-Abreu's representations associate low pitches (even one as 'salient' as that of the fall-rise) with a boundary which is not specified for tone.

This way of thinking about the mapping between pitch and tone requires us to acknowledge that a prosodic boundary empty of elements can have a (low) pitch

---

[6] 'A-licensing' - *autosegmental licensing* - which legitimises elements in a prosodic position contrasts with 'p-licensing' - *prosodic licensing* - which defines licensing relations between positions. Both are aspects of the fundamental principle of phonological licensing (Goldsmith 1990, Harris 1997, McCarthy & Prince 1986 *et al.*).

signature. In the case of English intonation, this association of low pitch and prosody allows the reduction of the combinatorial possibilities of the elemental inventory to the logical minimum (one privative element or none).

5.2.2 The conflation of [L] and [N]

Such reductionist ambitions are generally laudable, and one other particular enterprise bears directly on the relationship of [L] to phonetic characteristics. Nasukawa (1998) argues that in an analysis of Japanese (among other languages) it is possible to conflate the element [L] with the element [N]. In Japanese compounds, the phenomenon known as *rendaku* consists in the appearance of 'voicing' on an obstruent at the juncture of the two constituent morphemes. There is a constraint on this process (Lyman's Law) which prohibits its occurrence if the second term of the compound already contains a voiced obstruent. Nasukawa's explanation of these facts hinges on the identification of the active element in Japanese voiced obstruents as ***headed*** [$\underline{N}$] and in Japanese nasals as ***dependent*** [N].[7] Thus *rendaku* is the attachment of (floating) [$\underline{N}$] to the second term of a compound:

(2)

ori (*'fold'*) + kami (*'paper'*) = ori $\overset{[\underline{N}]}{\overset{|}{\text{gami}}}$

otoko (*'man, male'*) + kokoro (*'mind, heart'*) = otoko $\overset{[\underline{N}]}{\overset{|}{\text{gokoro}}}$

Such attachment is prohibited by a following voiced obstruent in the same domain because this would create a sequence of two headed [$\underline{N}$]s and this is regarded by Nasukawa as a violation of the OCP[8] (1998: 217).

---

[7] Earlier accounts of these phenomena by Nasukawa identified the opposite isomerism (1995, 1997): this issue is not relevant here.

[8] There are several versions of the Obligatory Contour Principle (OCP), which broadly converge on the following statement taken from Harris (1994: 172): 'At the melodic level, adjacent identical units are disfavoured'.

(3)

otoko (*'man, male'*) + koto**b**a (*'speech'*) = otoko**k**oto**b**a (*otoko**g**oto**b**a)

The dependent [N] in the nasal ([origa**mi**]) can contract an asymmetric relationship with the extant headed [N], and so this problem does not arise.

Japanese 'voiceless' (neutral) obstruents generally acquire voicing at the juncture of a compound. Some quirks in this process are successfully analysed in Nasukawa's account as a manifestation of the same OCP-based constraint. The potential sequence of two voiced obstruents that would result from [tob] (*'fly'*) + [ta] (*past*) (*[to**dda**]) cannot occur. A head-dependent relationship must obtain between two [N]s, and so [tonda] is the only possible output.

Nasukawa thus deposes [L] from the element inventory altogether. The typological symmetry displayed by the predicted and attested cross-linguistic distribution of [H] and [L] contrasts in oral stops (Harris 1994: 135) suggests that it may be more perspicuous to retain [L] for orthographic purposes if these two elements are to be conflated, so we will continue to do so here.

The identification of non-nuclear [L] with [N] could be empirically tested by looking at the facts of assimilation in other languages that possess fully voiced obstruents (and nasals). South of the European 'voicing' isogloss, there are several potential candidates. In southern French, for instance, the kind of assimilation in (4(a)) is attested (Price 1991: 126). This is contrasted in (4(b)) with apparently similar circumstances in (most dialects of) English, which do not display this phenomenon.

(4)

(a) *S. French:* avɛk + vu → avɛgvu 'with you'
(b) *English:* blæk + van // *blægvan 'black van'

Fully-voiced obstruents, as previously noted, are said to contain an active element [L]. The French assimilation, if it be phonological, therefore represents the appearance of this element in a position previously lacking laryngeal specification; there is no assimilation in English because English lenis obstruents have no [L]. Now the conflation of [L] and [N] broadly makes the prediction that in the circumstances exemplified in (5), where nasals meet neutral obstruents across an empty nucleus, assimilation will also be triggered.[9]

(5)

avɛk	+	mwa	→	avɛgmwa	'with me'
avɛk	+	nu	→	avɛgnu	'with us'

It is also predicted that liquids, like vowels, being innocent of laryngeal elements, should not behave like this. The stops in (6) should remain neutral.

(6)

avɛk	+	lɥi	→	avɛklɥi	'with him, it'
avɛk	+	ɛl	→	avɛkɛl	'with her, it'

The pronunciations in (5) are (informally) attested (Price 1991: 126), but it would be interesting to glean some hard evidence about these predictions.

The [L]/[N] conflation hypothesis, of course, leads ineluctably to the question of the contrastive representation of nasals and fully voiced obstruents language universally. One typical reductionist 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. In this system, headed expressions imply stricture and unheaded expressions do not. So both nasal and oral stops are headed, but with inverse positioning of elements, and fricatives

---

[9] 'Broadly', because once data relating to this prediction has been collected, it may be necessary to import considerations of isomerism (or some equivalent mechanism) into the analysis, as Nasukawa (op. cit.) has done for the Japanese examples.

are empty-headed expressions having both place and laryngeal dependents, as illustrated in (7).

(7)

[t]	[n]	[s]
$\underline{A}$	$\underline{L}$	(_)
H	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, pc).

(8)

[p^h]	[p]	[b]
$\underline{U}$	$\underline{U}$	$\underline{U}$
H		L

So in a language with fully voiced obstruents we can deduce the componential makeup of [v] to be [U, L, (_)], and this would be consistent with the representation of the voicing assimilations triggered by [v] in (4(a)) and the nasals in (5) as instantiations of the same process.

Finally, it is worth bearing in mind that this approach demands some modification to the traditional acoustic correlates of source elements cited in §1.4 (p.34) and listed in Appendix A (given here as (9)):

(9)

[H]	'high tone'	(in sonorant: rise in Fx) (in obstruent: high frequency aperiodic energy).
[L]	'low tone'	(in sonorant: fall in Fx) (in obstruent: continuous Fx).

These definitional attributes for [H] and [L] depend on a sonorant/obstruent bifurcation which cannot apply to these data. We may wish to propose instead that the defining environment for the differing correlates is prosodic. Acoustic changes in Fx are then mapped to source elements under nuclei, while phonation and nasality are associated with these elements in onsets. A consequence of this position is that 'syllabic' consonants become potentially componentially different from their onset-bound equivalents. In any language with syllabic consonants, full voicing and no lexical tone, [L] inheres in [l] and [n], but not in [l̩] or [n̩], and nuclear [L] can only logically show up in such a language as a post-lexical tone. There doesn't seem to be any particular objection to this from the present theoretical standpoint, for to presume upon the melodic identity of syllabic and non-syllabic sonorants is to maintain a misplaced trust in some phonemic concept.

For now, these speculations are left to future research. Some other ramifications of this variability of [L] (and [H]) correlates will be made explicit in §5.5.3 with respect to language universals, but we now return to our language-specific account of asymmetries in the tones of Yorùbá.

5.3 Yorùbá prosodic constituents

5.3.0 Bearing in mind that Fx contours may map to boundaries which are non-specified for tone (§5.2.1) and that [L] (and elements in general, by analogy) cannot be ***presumed*** to possess a particular phonetic signature (§5.2.2), we may take the next step in our examination of the asymmetries of Yorùbá tone by identifying the prosodic structures of the language. There are constraints on the appearance of certain melodic objects. For instance, Yorùbá has four nasalised vowels but nasalisation is not permitted in the first nucleus of an indigenous disyllable: [tɛsɛ̀] ('joy') is a permissible word, while *[tɛ̃sɛ̀] would not be.[10] The first stage in a discussion of these constraints in general, and tonal constraints in particular, is to achieve a clear idea of the domains in which they may be

---

[10] These constraints are further exemplified and discussed in §5.4.

deemed to operate.

5.3.1 Minima[11]

A CV syllable, by which we formally mean a filled nucleus licensing a filled onset, enjoys a special status in Yorùbá. It is the shape of the smallest root in the language, the canonical verb form. Now Yorùbá is possessed of an epenthetic vowel [i][12] (Pulleyblank 1988), but, as Ola (1996: 5) notes, no augmentation or epenthesis takes place in the imperative utterance of an intransitive verb. This is illustrated in (10).

(10)

wá	*to come*	wá	*come!*
lɔ	*to go*	lɔ	*go!*

Given that these verbs can be uttered in isolation, the only conclusion possible is that they are 'big enough' to function as independent prosodic words. There are no stand-alone lexical forms which have the shape V, VV[13] or VC[14]. Function morphemes *can* be lexically specified as V, VV, or VC, these forms being permitted presumably because they are not words but must be set within a larger prosodic structure in all phonological circumstances, whereupon the 'CV-minimality' constraint will certainly be obeyed. We discuss function words further in §5.3.2 and §5.4.

Further evidence for the primacy of CV is to be found in the truncation of loan verbs,

---

[11] This section owes a considerable debt to Ola (1993a) and (1996). These works are couched in Optimality-Theoretic terms, and Ola's assumptions about syllable structure do not entirely conform to our own. However, her proposals about the minimal size for prosodic objects (syllable and word) translate well into our model and give support to our eventual contention that Yorùbá prosody is parametrised with rightmost heads.

[12] A single example: 'doctor' is borrowed as [dɔ́kítà].

[13] There are thirteen VV nouns (with VCV alternants) in Yorùbá, in contrast with the vast majority of nouns which are VCV (Ola 1996). We will discuss nouns more fully later.

[14] This last prohibition is predicted anyway in a language which displays a general constraint against final empty nuclei (i.e. the OFF parameter setting).

where [h]-epenthesis will take place if necessary in order to maintain this shape, as exemplified in (11 (c)).

(11)

(a)	k͡páàsì	k͡pá		*to pass*
(b)	k͡pɔ́m̀bu	k͡pɔ́		*to pump*
(c)	ɛ́ńfı	hɛ́	*ɛ́	*to envy*

One final instance of the CV requirement can be shown in the permitted environments for intervocalic '[r]-deletion' in Yorùbá.[15] The conditions for this process are either that [r] must be surrounded by identical vowels, or that one of the vowels must be high (Akinlabi 1993). Examples are given in (12).

(12)

erùk͡pɛ̀	eèk͡pɛ̀	*sand*
òrìʃà	òòʃà	*god*
àdúrà	àdúà	*prayer*
èkuru	èkuu	*cooked ground white beans*

[r]-deletion cannot occur, however, if its application would result in a form wthat did not contain a legitimising CV sequence. The examples of [r]-deletion in (13) are impermissible.[16]

(13)

orí	*oí	*head*
àrá	*àá	*thunder*
oró	*oó	*pain*
ɔrɔ̀	*ɔɔ̀	*wealth*

---

[15] For a slightly fuller list and more examples, see Ola (1996: 5-8).

[16] All examples in (12) and (13) are from Ola (1996), based on Akinlabi (1993).

The existence of the canonical CV forms means that with regard to minimality, Yorùbá differs both from English, in which the minimal word has two nuclear positions (one may be empty), and from epenthesising languages such as Axininca Campa and Lardil (McCarthy & Prince 1986) which obligatorily augment words to contain two *filled* nuclei. This fact weighs against the notion that there exists in UG an inviolable principle such that the prosodic minimal word contains a binary foot (expressed in terms of morae, syllables, nuclei or whatever). In Ola's Optimality-Theoretic account the Yorùbá facts are explained by proposing a constraint - 'Properheadedness'[17] - which includes the requirement that a syllable be present in every word. Ola has argued (1993b) that in Yorùbá, only CV projects a syllable, and by ranking the Properheadedness constraint higher than other constraints demanding binarity, she derives the CV minimality condition. Quite independently of any characterisation put forward about 'syllable' or 'mora' - independently, in fact, of their existence - it is self-evident that for Yorùbá a single nuclear position satisfies word minimality, so long as this nucleus always *licenses* (in our terms) a filled onset.

Since we are not availing ourselves of the OT toolkit in the present discussion, how can we place this Yorùbá deviance amid the conspectus of languages that display minimal foot-binarity, in our own terms? One possibility is to preserve universality by simply proposing that Yorùbá has no feet, but rather that a single nucleus licensing a filled onset projects a (minimal) prosodic word. Yorùbá, which now lacks feet or branching nuclei, simply does *word-minimality* a different way. The language-specific well-formedness constraint has as its domain the (phonological) word, in which respect Yorùbá is just like English or Axininca Campa. Without the sub-word prosodic equipment possessed by these languages, however, Yorùbá must use the constituent structure at its disposal to signify minimality. It is not controversial that languages display eccentricity of this kind. If typologically more common phonological distinctions are not utilised, then others may be: the Yorùbá case is similar to the exploitation of relatively rich segmental inventories by languages that fail to utilise (say) lexical tone.

---

[17] Following Itô & Mester (1992).

The exoticism in this case specifically inheres in the use of onset-dominated material to do a prosodic job. A language-specific requirement for filled onsets is rather more familiar at the 'syllabic' level. North German will interpose a glottal stop at the beginning of any word that is otherwise vowel inital: thus <Ente> ('duck') is realised as ['ʔentə], <Atem> ('breath') as ['ʔaːtem], and <Opa> ('grandpa') as ['ʔoːpa]. Karok, an Amerindian Hokan language, is possessed of the same constraint. This language, however, also uses glottal stops to represent lexical material, as we can see from morphological prefixations represented by the fragment of data in (14), taken from Bright (1957).

(14)

	imperative	1st. singular	gloss
(a)	pasip	nipasip	*shoot*
(b)	ʔifik	niʔifik	*pick up*
(c)	ʔiʃkak	niʃkak	*jump*
(d)	ʔaxjar	nixjar	*fill*

In (14b), [ʔ] is lexical: affixation of [ni] takes place in an entirely parallel fashion to (a). In (c) and (d) the unaffixed forms are lexically vowel-initial and [ni] docks onto the word, swamping the vowel colour of the first nucleus (as is clear from (d)): initial [ʔ] in (c) and (d) is only there to deliver a phonologically permissible form for the imperative by filling the initial onset.

The claim made here is that this kind of 'filled onset' requirement may obtain at other levels of the prosodic hierarchy. Corroborative evidence for this is to be found in Takahashi (1994). Here, the stress assignment system of Aranda (central Australia) is explained in an OT framework by the interaction of five constraints. One of these is a reformalisation of the more familiar ONSET constraint - 'syllables must have onsets'. Takahashi understandably accuses this constraint of 'informality' since OT does not recognise an onset constituent. No friend to constituency himself,[18] he therefore restates the constraint in terms of licensing, thus: 'PREHEAD (Δ) - 'the head of a prosodic domain Δ must p-license a prehead' (1994: 494). In order for Takahashi to derive the stress-facts of Aranda, it is crucial that the violation of this constraint is not restricted to the level of the non-projected nucleus: 'Once higher prosodic structure is established, the violation of PREHEAD may be incurred by the same nucleus at different projections' (op. cit.: 494). The head of the (obligatory) disyllabic foot in Aranda, as Takahashi successfully shows, is obliged to license an onset to comply with PREHEAD.[19] Re-translating into standard OT, he extrapolates: 'this can be interpreted in terms of the constraint ONSET having arguments such as ONSET (s), ONSET (Ft) and ONSET (Wd)' (op. cit: 504).

The present contention, then, is that CV-primacy in Yorùbá be characterised as a requirement that the prosodic word licenses an onset. This minimal word is regularly the shape of the canonical verb-form. We can profitably contrast this with that of the noun - VCV. This is overwhelmingly the form of vernacular nominals, as exemplified in (15).[20]

(15)

ìlú	*town*	ɔmɔ	*child*	ɔdũ	*year*	
ɔ̀rɔ̀	*word*	àg͡bɛ̀	*farmer*	àwo	*plate*	
èdè	*language*	iʃé	*work*	odò	*river*	(etc, etc, ...)

Here the minimal prosodic word is augmented by a single vowel to its left. The initial 'syllable' of these nouns is degenerate in that such a form could never exist independently in the language. In Ola's (1993a) terms this position does not 'project a syllable node'. In terms of licensing, it seems reasonable to characterise the relationship between the two nuclei as follows:

---

[18] See Takahashi (1993).

[19] By OT, this obligation is of course violable under some circumstances.

[20] Examples from Bamgbose (1966).

> A nucleus lacking a filled onset may only be realised if it is licensed by a properly constituted minimal nuclear domain (i.e. a CV sequence).

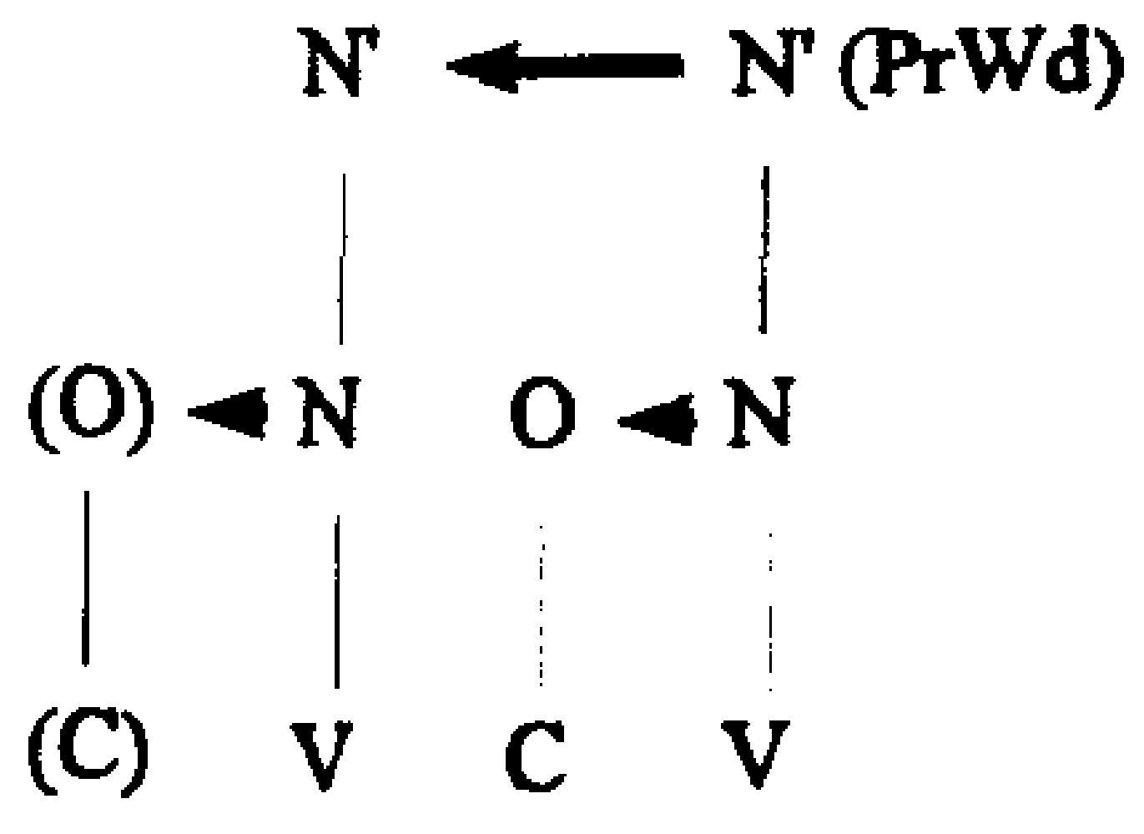
The head of the 'properly constituted minimal nuclear domain' is itself, of course, the nucleus. Notwithstanding the robustness of VCV as a nominal canon, a minority of indigenous[21] nouns (and other lexical items) exist which have the form CVCV, as exemplified in (16).

(16)

tayɔ̀	*joy*	bàbá	*father*	tɛsɛ̀	*foot*

so the inter-constituent prosodic licensing relations in the Yorùbá noun can so far be diagramatically represented as in (17).

(17)



In conformity with universals, onsets are licensed by the nuclei to their right. The right-headedness at the projected N' level (ProsodicWord in (17)) is the manifestation of the setting of a language-specific parameter. We know that the head of the Prosodic Word domain licenses an (obligatory) onset. We might also expect to see further manifestations of a greater autosegmental licensing potential from this position, and we will see that this prediction is ultimately borne out. A particular asymmetry that will be seen to assume some importance is that high tone may not be licensed in the leftmost dependent nuclear position. An account of this follows (in §5.4) after we round off our discussion of

---

[21] Foreign borrowings are discussed in §5.4.1, (p.150).

prosodic size with a look at maxima.

5.3.2 Maxima

We have seen (examples in (16), representation in (17) above) that the leftmost dependent nucleus in a word has the potential to optionally license a filled onset. The largest structure we have so far considered is CVCV, and on closer examination this turns out to be a maximal prosodic constituent in Yorùbá. It is, of course, possible to find 'words' in the language which are prosodically bigger than this (such as 'Yorùbá'), but once a word exceeds (C)VCV it is often possible to trace compositionality of meaning in its make-up, which gives an indication that the morphology is in fact analytic.

Ola (1996: 14) proposes that 'monomorphemic words are maximally two feet'. For us, however, this blurs the special status of (C)VCV - which is a disyllabic foot in Ola's account - and we have in any case so far proposed that there are no feet in this language, this assertion serving as at least a partial explanation of the CV primacy. Of the four examples of 'two-footed' words that Ola gives, at least two betray componentiality in her translations (quoted in (18)).

(18)

g͡baragada	*wide and open*	sɔ́rɔ́gɔ́dɔ́	*tall and slim*

The line of reasoning followed here has led to the assertion that there is no need to propose an independent foot level for Yorùbá. In this respect we have parted company with Ola's account. The characterisation in (17) (above) of (C)VCV as a 'ProsodicWord' is closer in spirit to the intuition expressed by the use of Siertsema's term 'basic word'.

> 'Basic words are the monosyllabic words of the structure CV and the disyllabic words of the structure (C)VCV not derived from them. These enter freely into larger words as morphemes...' (Siertsema 1959: 47).

Such agglutinations are very much the norm in Yorùbá, and are accompanied by the kinds of phonological processes which regularly accompany concatenative morphology. Vocalic assimilations and the loss of segments with or without the loss of tone are widely attested.

The two examples given in (19 (a) and (b)), which are taken from Bamgbose (1966) and Siertsema (1959) respectively, demonstrate that these processes can be traced within comparatively simple, as well as obviously complex, words.

(19)

(a)

bí *to give birth* + ɔmɔ *child* → bímɔ *to have a baby*

(b)

	à	+	kó	+	so	→	àkóso
	(noun-forming prefix)		*together*		*to bind*		*control, government*
then							
	a	+	lí	+	àkóso	→	aláàkóso
	(noun-forming prefix)		*to have*		*control, government*		*boss, president*

The thrust of Siertsema's argument was to disabuse linguists of the notion that a vocalic length contrast obtained in Yorùbá, a battle long ago won. The boundaries around the 'basic word' are obviously crucial to her idea, as is the juxtaposition of identical vowels bearing different tones: all 'long' vowels are in fact sequences of two short vowels That being so, each will head its own nuclear domain. The existence of a strong prosodic juncture at (at least) the edges of a maximal (C)VCV, and one that survives morphological concatenations, will have other consequences for our discussion of Yorùbá, which we take up in §5.4.2 (below).

To summarise this section, it turns out from a consideration of the constraints on word-size in Yorùbá that they are unusual, but not impenetrable: there is a minimum CV and a maximum (C)VCV.

5.4 The form and content of Yorùbá nuclear domains

5.4.0 So far in this discussion, we have delimited the size of Yorùbá words, but there has only been a passing mention of sub-word formatives. There are non-stand-alone objects in Yorùbá that are V-sized: these may only be realised if they are legitimised by joining a structure which conforms to the minimal word-size constraint (contains a CV sequence). We will see that these objects do not conform to melodic constraints on dependent domains, and so must constitute a domain of their own: for this reason we must consider the question of the potential existence of a sub-word prosodic constituent. Before we do, the constraints themselves will be enumerated.

5.4.1 Constraints in dependent nuclei

Asymmetries in the Yorùbá noun canon show constraints in the permitted melodic content of dependent nuclei. This right headed (C)VCV structure allows a full set of melodic possibilities in the head nucleus, but its leftward licensee abides by the following three conditions:

(20)

(a) No high tone
(b) No nasalised vowels
(c) No high back vowels.

Before proceeding to consider the particulars of the facts in (20), it ought to be mentioned that it is entirely usual for a language to exhibit low licensing capacity within a relatively weak nuclear domain. Hollenbach's (1984) dissertation concerning Copala Trique, an Oto-Manguean language from Mexico, shows that in (licensed) final syllables there are no stresses, no source contrasts, no tonal specifications and no nasalised vowels. All of these possibilities are only available within strong nuclear domains. This kind of asymmetry can be formalised by adopting the notion of 'licensing inheritance' (Harris 1994, 1997) whereby positions at a greater remove from the head of a domain are progressively less able to a-license melodic material. The basic principle of licensing inheritance in a melodic context is that ' a licensed position inherits its a-licensing potential from its licensor' (Harris 1997: 340).

Ola (1993a: 56-58) extends the constraints in (20) to all V-initial words in the language: 'vowel-initial high-tone words are completely unattested' (op. cit.: 56). Presumably she considers, as do we, non-stand-alone forms such as pronomial [ó] ('(s)he') as something less than a word (at least phonologically), in which case her argument that an extension of these constraints is to be seen in prefixation (op. cit.: 57) is automatically gainsaid. As previously noted, Ola goes on to characterise the asymmetric behaviour of V and CV as indicative of the fact that V is extrasyllabic: specifically that the 'onset-less V is an unsyllabified mora, that is, a mora that is not dominated by syllable structure in the phonology' (op. cit.: 60). The ramifications of Ola's proposal that moraic licensing takes place outside the syllable seem unappetisingly rich, but what is in any case not delivered by her account, as it is outside the scope of its aims, is a formalisation of the interaction of the V/CV asymmetry with the facts which reflect the particular constraints in (20).

Ola's formalisms and our own diverge irretreivably when it comes to a consideration of the 'syllable'. While it is a genuine category in her system, we use the term informally for a nuclear domain, and as is usual in GP, no syllable node is present in prosodic structure. To recapitulate our position so far, the pervasive importance of (C)VCV in Yorùbá has led to its characterisation (in (18) (above)) as a maximal prosodic word, wherein neither syllables nor feet intervene between the nuclear level and the word level.

We are now faced with the hitherto unposed question of how we relate the constraints in (20) with the (autosegmental) licensing potential within the domain word in a simple way. It is patently obvious that there are restrictions, in our terms, on the elemental content of the licensed nucleus. Using the traditional element inventory (§1.4, p.34), we would have to have imposed a ban on [N], [H] (but not [L]) and simplex [U], hardly a felicitous natural class.

If the mapping of [L] to nasality in Japanese (§5.2.2, p.134) is accepted and extended to Yorùbá, we can improve on this. The lack of high tone in licensed nuclei (20 (a)) remains as a proscription against [H]. The lack of nasality (20 (b)) becomes a proscription against [L]. It is now possible to generalise from these two constraints:

(21) Source elements cannot be a-licensed under a licensed nucleus.

In fact, it constitutes a neat piece of evidence for the NLT hypothesis that in Yorùbá licensed nuclei parallel constraints exist against high tone and nasality, rather than against high and low tone. The former pair, not the latter, are seen to be phonologically congeneric under these circumstances.

The proscription against high back vowels (20(c)) will have to remain: further generalisation is not possible. Yorùbá has seven oral vowels, represented in (22):

(22)

i u
e o
ɛ ɔ
a

There are two autosegmental tiers relevant to vowels: the [A]-tier and the 'colour' (fused [I] and [U]) tier. In addition to the simplex peripherals, the language has two parallel pairs of mid-vowels, which are isomerically distinct. The lower member of each pair is unheaded by dint of the general constraints against [A]-headedness (§3.5, p.81), and the higher member of each pair 'colour'-headed.[22] The block on [u] in licensed nuclei cannot therefore be expressed as a ban on headed, [U]-headed or 'colour'-headed expressions: there are plenty of words like [iʃɛ́] (*work*) and [ɔ̀fɔ̀] (*loss*) which show this notion to be

---

[22] This statement is more concordant with the Rennison (1990) model, wherein headedness is the property of a tier, than with other analyses which represent headedness as the property of an element.

counterfactual. The only conclusion is that the melodic deprivation of the licensed nuclei rather mysteriously leads to this ban on simplex [U].

5.4.2 Sub-word prosodic domains

The proposal that low pitch, or a fall, does not cue an element in Yorùbá can be argued to underpin some further asymmetries in the tonology of the language. Before proceeding to enumerate any benefits that it creates, though, we need to solve an immediate representational problem that it generates. How do we formally account for minimal pairs like [iʃɛ́] (*work*) and [ìʃɛ́] (*poverty*)? If there is no melodic element in the licensed nuclei in either word, in what does the low-mid distinction inhere?

A clue to the answer may be found in the 'Yorùbisation' of foreign nouns. Two examples are given in (23).

(23)

tíʃà	'teacher'	kólà	'collar'

This is the universally attested Yorùbá form for borrowed English trochees. Note that they are impossible as single phonological words, because of the high tone in the first nucleus. We can draw a parallel between these and the 'total' numbers in (24), which contain transparently analytic morphology.

(24) méɟì, mɛ́ta, mɛ́rĩ, márũ, mɛ́fà, méɟè, mɛ́ɟɔ, mɛ́sa... '2,3,4,5,6,7,8,9.....'

All of the examples in (24) are compounded of a 'total' prefix and a 'basic' number. 'Basic' numbers are legitimate VCV words bearing 'low tone' on the dependent nucleus. One example of the concatenative process will suffice: it is the same in every case.

(25)

mɴ́ + ɛ̀ta → mɛ́ta

The N marked for high tone here indicates an empty nucleus: though the first vowel of the number persists into the compound, it is the prefixal high tone which will swamp the original low. The survival of high tone in these circumstances is very much the norm in west African languages.[23]

Unlike genuine CVCV words, such as [tayɔ̀] (*joy*) and [bàbá] (*father*), the resultant compounds apparently violate the proscription in (21) against source elements in the dependent nuclear position, just as the loan words in (23) do. A solution is available if we put aside the morphological differences between the examples in (23) and those in (24), and propose that they share the property of being phonologically analytic. As a single domain, the first 'syllable' of [tíʃà] is on a prosodic par with the verb [kí] (*greet*): it is an unlicensed head and is free to a-license [H]. Total numbers and borrowings demand that the first nucleus in each case be head of a domain: given right-to-left directionality (cf. the discussion in §5.3.1 and §5.3.2 above), and that these disyllabic objects are prosodic words, we must now include sub-word prosodic structure in our analysis. This is the proposal:

> **SW** There are sub-word boundaries in Yorùbá whose existence can be deduced from sequences which would otherwise be phonologically illegal. From hereon, the constituent contained within such a boundary will be referred to as Subword (abbreviated as SW).

There are only two logical possiblities for the items in (23) and (24): [[CV]CV] and [[CV][CV]], both involving an independent domain for the first nucleus. Can either or both of these structures be motivated? Since it is a commonplace (but not a universal) in language that prosodic boundaries are signalled by low pitch or a fall, we might expect that the 'extra' domain we have mooted be mapped in such a way. This would give us a parallel (though at a different prosodic level) with the non-tonally-specified English intonational boundaries discussed in §5.2.1 (p.132), which are also cued by low pitch.

---

[23] See the 'tone hierarchy' in §5.1, (p.130), and the examples from Ogori in §5.5.1, (p.164).

What was previously considered to be a 'low tone' is now the presence of SW unspecified for a laryngeal element. [[CV][CV]] would then have to have lower pitch in the second nuclear domain than does [[CV]CV], deriving 'low' versus 'mid' tone. The parallel between the morphologically analytic numbers which betray their phonological analyticity by bearing a leftmost high tone, and the borrowings, which are now also to be considered phonologically analytic, is diagrammatically represented in (26).

(26)

(a) *'Total' numbers* [[mé][ɟɪ]] [[mɛ́][fa]]

(b) *Borrowed trochees* [[tí][ʃa]] [[kó][la]]

The orthographic low tone mark has been abandoned in (26) as it is now redundant. The numbers with 'low tone' on the second nucleus have a shape exactly coincident with the trochees: those with 'mid tone' are exemplified in (27),

(27)

[[mɛ́]ta] [[mɛ́]ɟɔ]

while the nasalised vowels can now be analysed as having an [L] component. It is, on this analysis, the contrastive prosodic constituency of these forms which maps to differing pitch characteristics.

To derive the parallel structures in (26), we have explicitly ignored morphology. This step requires no special justification, as phonological domains are universally *potentially* autonomous of morphosyntax. Parallels in phonological analyticity can be discerned at quite different levels of morphosyntactic structure. In English, root-level morphology and underived forms display identical phonotactic constraints, which are relaxed equally across sentence junctures and 'word'-level morphological boundaries (Harris 1994: 50-52). Within compound words, the existence of an accidental juxtaposition gives an unambiguous indication of such a boundary, and bracketing paradoxes can then provide clear instances of the mismatch between morphology and phonology.

A classic English example runs as follows. Prefix [un] is a member of the class of prefixes which maintain independent phonological wordhood in a compound; this delivers the otherwise impossible geminate in [ʌn'nætʃrəl]. For this reason, the word 'ungrammaticality' must be phonologically analysed as [[un][grammaticality]]. But morphological *un-* is a nominalising prefix: it does not attach to nouns. The efficacy of the Orwellian stylistic grotesque 'un-person' only serves to underline this. The morphological analysis of [[un[grammatical]]ity] is therefore not coincident with the phonological. Bracketing paradoxes are mortal wounds to a system (such as 'Lexical Phonology') which argues for coincident morphophonological levels: thus '...it is...firmly established that morphological and phonological (cyclic) structure need not be isomorphic' (Kaisse & Hargus 1993).

A closer genealogical parallel to the proposed lack of coincidence of phonology and morphology in Yorùbá can be found in Zsiga's conclusion about Igbo compounds: 'The two elements of a compound ... count as two constituents prosodically, but a single constituent syntactically' (1992: 133). Zsiga shows the domain of vowel (ATR) harmony in Igbo to be the (prosodic) word. Harmony does not apply between the two elements of a compound, so they must remain prosodically discrete. However, the morphological derivation of these forms marks them as a single (word) domain: inflectional and extensional affixes may only attach at the edges of such a compound, never between the elements.[24]

---

[24] For the sake of completeness, it is worth while briefly laying out the view taken in this discussion on the relationship between the formatives of *morphology* and *syntax*. Apart from the occupancy by *morpheme* and *word* of different structural levels, there seems nothing to choose between them. There follows an extract from an essay (Harrison 1994) which presents the position adopted here, a position which simply conforms to the majority of the literature on this subject (e.g. Baker 1988, Lieber 1992).

> 'What makes <kind> a word, and <-ness> an affix? The two possibilities that immediately occur are that <kind> can have an independent existence, and that it has 'lexical' content. This latter possibility rapidly disappears unless we wish to debar <the> and all other functionals from wordhood. And in what sense can <kind> be more independent? It requires an English sentential structure to be of any semantic or syntactic use, and is constrained in its distribution in that structure, whichever syntactic theory is adopted. It must appear in an adjectival slot as surely as <-ness> must appear in a suffixal slot. While it is true that <-ness> is more constrained in its distribution than <kind>, this is a relative, not an absolute, distinction ... We have made no progress at all in separating <kind> from <-ness>: they simply appear to be constituents of different categories. The 'morpheme' has become indistinguishable from the 'word'.'

We have proposed that SW is represented at a projected level but not as high as the word. To return to the distinction between [iʃɛ́] (*work*) and [ìʃɛ́] (*poverty*), we may use SW to align these problematic *mid/low* contrasts with the rest of the story. Yorùbá SW, like stress in other languages, may be a part of the lexical address of a morpheme, and may be deemed to operate at the same prosodic level, that which is more generally characterised as 'foot'. Thus there is a parallel between [iʃɛ́]/[ìʃɛ́] and a stress minimal pair such as the modern Greek [pɔ'li] (*very*) / ['pɔli] (*city*). More discussion from a perceptual standpoint about the exoticism, or otherwise, of the proposal that SW exists will follow in §5.6 (p.169), when we propose that in some sense 'foot' and 'SW' may be considered as mirror images. Meanwhile, the relevant phonological aspects of the two Yorùbá words, on this account, can be diagramatically represented as in (28 (a & b)).

(28)

(a) [iʃɛ́]

[H]
|
[iʃɛ]

(b) [ìʃɛ́]

[H]
|
[[i][ʃɛ]]

Note that there is no obligation on [H] to spread within the word domain any more than there is on any other element.

A further tonal asymmetry may be cited as circumstantial evidence to undermine any notion that high and low tone are congeneric. When two words are conjoined, the juncture is sometimes marked by the appearance of an independent morpheme comprising a (floating) high tone. There is no analogous participation in such a process by low tone. An example (from Bamgbose 1966) is given in (29).

(29)

ìyɛn 'that' + ʃòro → ìyɛ́n ʃòro 'that is difficult'

This particular distributional asymmetry is explained if we acknowledge that high tone is possessed of [H] as an active element, and look elsewhere for a phonological analysis of low.

Before a further consideration of sub-word constituents, the aspects of Yorùbá prosody considered so far are recapitulated in (30).

(30)

(a) Word size constraints operate at the level of prosodic word: the minimum is CV, while the maximum is (C)VCV.

(b) Words are parametrised as right-headed.

(c) At the projected level normally relevant in languages for stress-feet, SW either *is* or *is not* present in a given context.

(d) Examples have been cited where SW functions to prevent the formation of impossible words.

(e) SW has an acoustic cue: it is typically low-pitch or a fall (it is in fact all those phonetic reflexes previously associated with low tone).

5.4.3 Enigmatic alternations

Several authors have previously proposed that tones cannot be regarded as of equal phonological status in Yorùbá, though without invoking any particular formalisms, and certainly without recourse to the ones under discussion here. Stahlke (1974) cites as evidence for tonal asymmetry the productive verbal paradigms of the language, and the behaviour of tonally-specified clitics within these paradigms.

Using the verbal root [lo] (*to go*) as an illustration, Stahlke shows that the presence of a clitic lexically marked for high tone will result in a predictable alternation in a preceding 'mid tone' pronominal between mid and low tone, the high tone selecting a preceding low. This is illustrated in (31), where the progressive marker [ń] and the future marker [á] are prefixed onto the root following the first and second person singular and plural pronominals.

(31)

past	*progressive*	*future*	
mo lo	mò ń lo	mò á lo	*I*...
o lo	ò ń lo	ò á lo	*you* (sing.)...
a lo	à ń lo	à á lo	*we*...
e lo	è ń lo	è á lo	*you*(pl.)...

No such alternation is observed if the pronominal is itself marked for high tone: in the third person verb forms in (32) the 'robustness' of high tone is evident.[25]

(32)

past	*progressive*	*future*	
ó lo	ó ń lo	ó á lo	*(s)he* ...
wó(n) lo	wó(n) ń lo	wó(n) á lo	*they*...

The net result of these observations is that, while high-tone pronominals are always high-tone, there is a representational issue to be clarified with respect to the alternating forms. Stahlke proposes the tonal distinction in Yorùbá to be a contrast 'between just high and a non-high tone, rather than between high, mid and low tones' (Stahlke 1974: 139), which at the very least obliquely concurs with the NLT hypothesis under discussion here. But exactly *why* a high tone should select a preceding low remains to date an open question. The dissimilatory nature of this process hardly allows it to pass muster as a generalisable phenomenon, unless it were to be categorised as an instance of some kind of polarity.[26] However, the attractiveness of this particular idea wanes after a consideration of Yorùbá words in general. There appears to be no distributional regularity to favour a word like [ìlú] (*town*) over [ilé] (*house*). From a random collection of 73 VCV nouns with final high tone, 40 have an initial mid and 33 an initial low.

If the alternation is the result of the operation of a phonological process, it must be right-to-left directional, in conformity with the licensing parameters we have discussed. In the

---

[25] Parenthesised (n) indicates that the preceding vowel is nasalised.

[26] This consideration was raised by Glyne Piggott (pc).

examples in (31) above, it looks as if the tonal specification of the rightmost (tense-marking) clitic affects the shape of its leftward neighbour. If an SPE-type formalism were to be invoked to describe the process in the context of two phonological tones, but at least allowing sufficient non-linearity into the model such that nuclei are local at a projected level, the rule would look something like (33).

(33)

$$\emptyset \rightarrow L / \_\_ \; H$$

This is not a well-motivated rule.[27] If the juxtaposition of Ø and H triggered a dissimilative insertion of L, then under identical morphological circumstances, it might be expected that the juxtaposition of H and H gave rise to a similar effect. The forms in (32) show that this does not happen.

If we try to use more contemporary models, and adopt Stahlke's proposal that only one tone exists, things just get worse. It is impossible to see how any activity on the part of [H] could 'percolate' to the left and lead to anything other than tone-copying (or spreading). Dissimilation would appear to be out of the question.

It is therefore necessary to look elsewhere for an explanation of these alternations. The following facts should be incorporated in any account:

(i) Once concatenation has taken place, all and only the past tense forms in (31) form legitimate single words of Yorùbá, exemplified by [o lo] and [mo lo].

(ii) The past tense forms in (32) must be phonologically analytic because the leftmost high tone must be in the head position of a phonological domain.

---

[27] The elegantly simple shape of this unlikely rule is, of course, one of the weaknesses of the rule system as a whole.

(iii) All the progressive and future forms must be analytic because they exceed the maximum word size.

(iv) Both of the tense markers and four of the six pronominals in (31) and (32) are subminimal because they do not meet the 'filled onset' requirement for a word (see §5.3.1, pp. 139-145).

In the previous section, SW was invoked as a phonological strategy in Yorùbá to prevent the formation of impossible sequences. Now if we assume that the pronominals are formatives, rather than words (which is what most of them are obliged to be by minimality) and that [lo] subcategorises in the past tense for a pronominal formative, and in the others for a [pronominal + tense marker] formative, we can achieve some explanation for the alternations. Subcategorisation of this order for the progressive and future forms is supported by the following observation: at no stage in derivation could the sequence [high+mid] (e.g. [á lo]) be legitimate (by licensing constraints), while the sequence [low+high] (e.g. [ò á]) passes muster in this respect, even though it violates Yorùba minimality.

Proposing this derivation means that the progressive and future pronominalised verbs have integrity as a unit. They must therefore be marked out as exceptional as they contain three, rather than two, nuclear domains. The juxtapositions of two subminimal components suggest that these are unusual phonological circumstances.

Before making the prosodic shape of these concatenations explicit, it will be useful to review the tonal content of the empty-onset formatives that have been predicted by the present account.

(34)

$$\mathrm{N} \qquad\qquad [\mathrm{N}] \qquad\qquad \begin{array}{c}[\mathrm{H}]\\ | \\ {[\mathrm{N}]}\end{array} \qquad\qquad \begin{array}{c}{}^{*}[\mathrm{H}]\\ | \\ \mathrm{N}\end{array}$$

The first three representations here translate as mid, low and high-toned nuclei, while the fourth is debarred, as the laryngeal element is only permitted within SW.

The lexical forms of the pronominal prefixes have 'mid-tone', i.e are not laryngeally specified and not contained within SW. In the past tense paradigm, these combine with the verb to deliver a legitimate prosodic word. However, when the pronominal and the tense marker are juxtaposed, the resultant form is still less than a word (for example, [o[á]]). A different, and more exotic, role can now be seen to be taken by SW; the presence of this phonological domain reflects the subminimality of this structure, and hence the low pitch acoustic cue is realised. The representations in (35) exemplify these forms, with the SW domain boundaries marked with large, emboldened brackets.

(35)

(a)

[o lo] *you went*

(b)

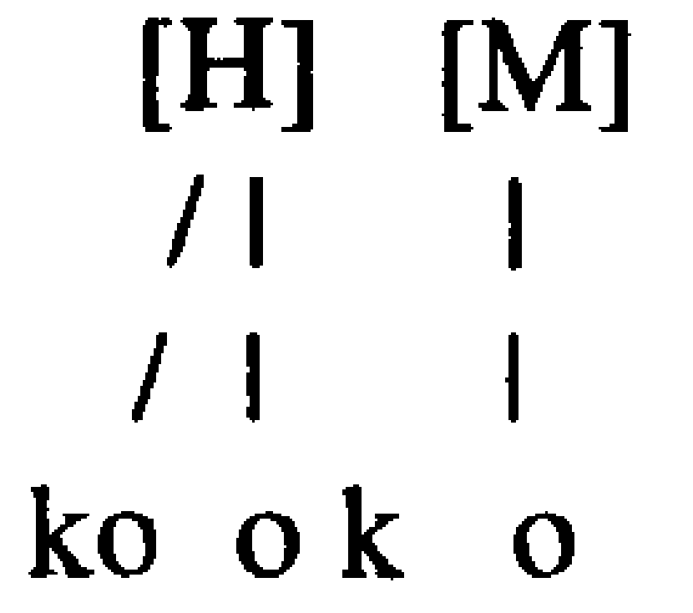
```
  [H]
   |
[[o[a]] lo]    (you (sing.) will go)
```

Even though there is no lexical significance in the signature of SW in the progressive and future forms in (31), it has been possible, following the arguments in this chapter, to discern a phonological motive. These forms, unlike the past tense, exceed the maximal word size *and* juxtapose two subminimal objects. The second of these, in each case, is lexically specified for at least one source element and so must be a head. Sub-word prosody is proposed to legitimise these structures in a parallel fashion to its operation in the borrowed trochees.

The drift of this discussion is reductionist. While it is true that by propounding the notion of SW we have added an extra level to our original prosodic frame (see (17) in §5.3.1, p.144), Yorùbá remains within the conspectus of languages in possessing a single intervening level of phonological structure between the nuclear domain and the prosodic word. This permits a reanalysis of the melodic componentiality of the language using an impoverished elemental inventory, and allows us to begin to account for the tonal asymmetries of the language.

It is perhaps worth comparing this endeavour with work in which the tonal specifications available in Yorùbá are actually *enriched* from the logical baseline of two. Oyètádé (1988) proposes three equal-status representations for lexical tones: [H], [M], and [L]. He argues that a phonologically active [M] is needed to block the spread of [H] in a word like [kóóko] (*weed*). His representation is given in (36) (from Oyètádé 1988):

(36)



Oyètádé does not delimit word boundaries in his illustration: we may want to assume from our proposals about maxima and right-headedness that this is a [[CV][VCV]] structure. This would, if true, support Oyètádé's hypothesis, given some kind of bracket erasure after concatenation of the components. There is a crucial objection to this. A now-familiar constraint on Yorùbá [VCV] words is that the leftmost (dependent) nucleus may not bear high tone (as the second nucleus does in [kóóko]). It therefore looks as if, at the lexical level, the second nucleus is isomorphic with the third person singular pronomial -[ó]- and the structure is optimally analysed as prosodically triplex: [[CV][V]CV]. Given that a boundary (SW) intervenes between the second nucleus and the following CV, we have a reason for the tonal shape of the word, given in (37).[28]

---

[28] And, of course, no 'bracket erasure'.

(37)

[H] [H]
| |
| |
[[ko][o]ko]

With respect to the redundancy of [M], we might also remember that words are right-headed. There is no reason why [H] should participate in any rightward activity, even without reference to the siting of boundaries. Further to this, no grounds exist to suggest even the leftward spread of a laryngeal element in Yorùbá.

The juxtaposition of the two high tones in [[kó][ó]ko] can now be seen to be as phonologically accidental as the English geminate across the prosodic boundary in '[un[natural]]'.[29] This analysis seems as plausible as Oyètádé's, and refloats Yorùbá in the mainstream of its dialect continuum, where

high >>> low >>> mid,

this hierarchy being far too well-attested to permit an analysis of three equal-status tones.

It is worth emphasising the situation of SW at a level lower than the prosodic word. If it were proposed that low pitch must always mark *word* boundaries, we would arrive at the plainly counterfactual prediction that there should never be any sequence of more than two contiguous mid-tones. Connell & Ladd deliberately created like-tone sentences for testing, including the following example containing a sequence of ten mid-tones (1990: 8).

(38)

omo wɔ̃ ni e lo fi se oko *It is their son that you marry.*

Though there does appear to be a general constraint against sequences of more than two identical tones in longer stretches of Yorùbá, based on a count of like-tone sequences in

---

[29] See §5.4.2 (p.153) for a brief account of English morphophonology.

the 1,367 syllable text given in Appendix V of Bamgbose (1966: 167-171)[30], this is probably accounted for by a general psychological (and *not* linguistically modular) human preference, and (38) can be presumed to be an entirely natural utterance. There is no reason why a sequence of mid-tones (i.e. nothing) should need fixing like the impermissible words in the verbal paradigm in (31) (p.156). The only subminimal object in (38) -[e]- is at a major clause boundary and so may be a different kettle of fish prosodically from the objects we have been considering. In any case, the following partitioning of this utterance into prosodic words is quite legitimate.

(39)

[omo] [wɔ̃] [ni] [e lo] [fi] [se] [oko]

We may wish to reconsider the rather arcane remark made by Siertsema, cited in Bakare (1995): 'short syllabic utterances ... constitute the most frequent structure in the Yorùbá language' (Bakare 1995: 54). The simplest formal interpretation of this statement is that the language is comparatively richly scattered with prosodic boundaries. This interpretation was advanced in §4.2.1 (p.88) and the assertion has now been supported with an analysis of the verb and noun canons.

In the interests of completeness it would be helpful to align these proposals with the Yorùbá connected-speech local effects detailed in Connell & Ladd (1990), particularly the two 'unexpected' effects that result from the juxtaposition of high and low tone. The most robust of these is H-raising before L (op. cit.:17-18). If this were interpreted as the presence of [L] influencing the phonetic character of [H], it would run counter to everything we have said. This is, of course, not the only hypothesis possible, nor do Connell & Ladd advance this or any other specifically phonological proposal. The other local effect which they note is that L is lowered between two H's. These two phenomena *may* be linked: 'it might appear that in HL sequences the contrast between them is exaggerated in both directions by raising the H and lowering the L' (op. cit.: 19). But they

---

[30] A rough correlation between prediction and attestation gives for two-tones-alike strings 408 against 386, but for three-tones-alike, 136 against 64.

also raise another possible interpretation of the second of these effects: it is potentially due to a suprasegmental effect. 'This lowering of L, if it can be shown to occur in any HL sequence... is (possibly)[31] not a local effect at all, but the manifestation of downstep' (op. cit.:19). Either way we get no closer to a phonological characterisation of the objects underpinning these manifestations. In the final analysis, therefore, it is impossible to say very much that could be of use in relating the ideas presented here to Connell & Ladd's interesting results, since they find it impossible to characterise those results in phonological terms.

5.5 Typological context: other languages and universality

5.5.0 In the introduction to this chapter, it was stated that the asymmetric tonology of Yorùbá in no way marked the language out as unique. To help set the language in a wider context, some aspects of the tonology of two related languages are now described. One of these, like Yorùbá, has three surface tone heights, while the other has four. The superimposition of the NLT hypothesis onto these data turns out to make phonological sense.

5.5.1 Ogori noun compounds

Ogori is also a member of the Benue-Congoid phylum, though it differs enough from its linguistic neighbours to be considered an isolate. It is spoken in Kwara state in central Nigeria, north of Benin City.

In this language, if a noun/adjective combination should result in the juxtaposition of two vowels, the final vowel of the noun will be elided. It is evident from the examples in (40) that this is a process which targets prosody, as it proceeds independently of the colour of

---

[31] Parenthesis added.

the vowel.[32]

(40)

ìgìlà	*yam*	òkeke	*small*	ìgìlòkeke	*small yam*
úwú	*dog*	ɔ̀nɛbɛ	*that*	úwɔ́nɛbɛ	*that dog*

The tone that is associated to the final nominal vowel does not necessarily suffer the same elision. In the second example given in (40), it is evident that the high tone from the noun has eclipsed the adjective-initial low tone. This is entirely regular in the language, as illustrated in (41).

(41)

ɛsá	*cloth*	ɔ̀rĩrĩ	*black*	ɛsɔ́rĩrĩ	*blackcloth*
ɔ̀tɛ́lɛ́	*pot*	ɔ̀kɛ̃ka	*big*	ɔ̀tɛ́lɔ́kɛ̃ka	*big pot*

So far, it is still logically possible to propose that this process is dependent on prosodic position, rather than the nature of the tone, i.e. that it is the first tone of the adjective that is subject to delinking. The example in (42) shows that this is not so, but that high tone will supercede low, regardless of directionality.

(42)

ɔ̀bɛ̀lɛ̀	*mat*	óbòrò	*good*	ɔ̀bɛ̀lóbòrò	*good mat*

Further, instances such as (43) demonstrate the action of the hierarchical relationship of low and mid tones: in parallel circumstances to the examples involving high and low tones, low tone will survive at the expense of mid.

(43)

ɔdɔ	*rat*	ɔ̀nɛ	*this*	ɔdɔ̀nɛ	*this rat*

---

[32] All the Ogori data in this section is ultimately from Chumbow (1982): these particular examples are cited in Halle and Clements (1983).

In the terms we have developed in this chapter, all of these facts can be seen as an instance of structure preservation. If there is a boundary at the edge of SW in a component it is also there in the concatenated form, and if a nucleus is tonally-specified (has [H] componential), this specification also persists. Sample representations are given in (44).

(44)

[ɔdɔ]	+	[[ɔ]nɛ]	→	[ɔ[dɔne]]
		[H]		[H]
		|		|
[[ɔ][bɛ][lɛ]]	+	[[ɔ][bɔ][rɔ]]	→	[[ɔ][bɛ][[lɔ][bɔ][rɔ]]]

5.5.2 Igede tone heights

To further contextualise Yorùbá tonology, we look at how the NLT hypothesis may be applied to a language possessing more than three tone heights. Igede, a Benue-Congoid language from eastern Nigeria has four surface tone heights, which have previously been transcribed as H, M, 'M, and L. Intertonal relationships exist between the pairs HM, H'M, M'M, and 'ML (Stahlke 1977, Hyman 1986). Analyses using feature-based models have proved less than satisfactory in capturing these relationships. For instance, H and 'M do not appear to share any features, so any featural link between them looks fairly tenuous. Hyman (1986: 119) highlights this problem, and goes on to invoke lack of specificity, laying out the exhaustive logical possibilities offered by an analysis using two underlying tones, as given in (45).[33]

(45)

L	'M	M	H
[L]	[ ]	[H L]	[H]

It is blindingly obvious that characterisations of this type can be accommodated easily

---

[33] This is one of two alternative analyses offered by Hyman. This one is taken from Clements (1983). The other follows Yip (1980) and Pulleyblank (1986). Hyman finds them both equally appropriate: the Clements account has been chosen here as it is closer in spirit to our present theoretical framework.

within a theory of privative elements, with 'M, though melodically unspecified, nevertheless receiving 'phonetic interpretation'.[34]

The advantages of this decomposition are then exemplified. The given instance is that in Igede an object clitic is realised with M tone after H or M, but with 'M after 'M or L. Hyman states simply that the clitic starts toneless and 'acquires the tonal specification of M ... after primary H' (Hyman 1986: 119). This type of process is, on this account, said to be the result of spreading. We might now wish to ask why it is only the presence of [H], and not that of [L], that drives the alternation, and why yet again we see an asymmetry of this fashion. It is also worth mentioning that in Hyman's general framework, which is rooted in feature theory, H is regarded as the positive value of single tonal feature ([+T]) and L as its negative ([-T]). It is hard to see how M could be specified as both plus and minus [T], unless these specifications occupied different tiers: the existence of primary and secondary tiers for tonal representations is also part of his account.

If we apply NLT to Igede, we straightaway encounter the problem of how to account for the difference between L and 'M without [L]. By analogy with Yorùbá (and English intonation), the prosodic distinctions in (46) may be proposed:

(46)

L	'M	M	H
[[ ]]	[ ]	[H]	[[H]]

and we can then account for the asymmetric behaviour of the object clitics. The deposal of [L] from the array of tones in (45) allows a straightforward account of this process as the spreading (in Hyman's terms) of the only available source element.

We will not pursue the specifics of Igede or Ogori further, but cite these instances as circumstantial evidence bearing on the assertion that H and L cannot be represented as

---

[34] The lack of continuity in the representations of Igede tone, *viz.* [H] >> [HL] >> Ø >> [L], where >> means 'is higher than' is not a problem, as there is no requirement that height should correlate with complexity.

equal-status objects in west African languages. Later work than Hyman's has, as we noted in chapter 2, successfully analysed southern African languages as being possessed of [H] only (see, for instance, Demuth 1993), but from the Igede and Ogori facts so far presented in this chapter, as well as from Yorùbá itself, it is possible to conclude that the logically motivated [H] / [L] / Ø analysis of west African nuclear tones is also unsafe.

One other point made by Hyman demands our attention:

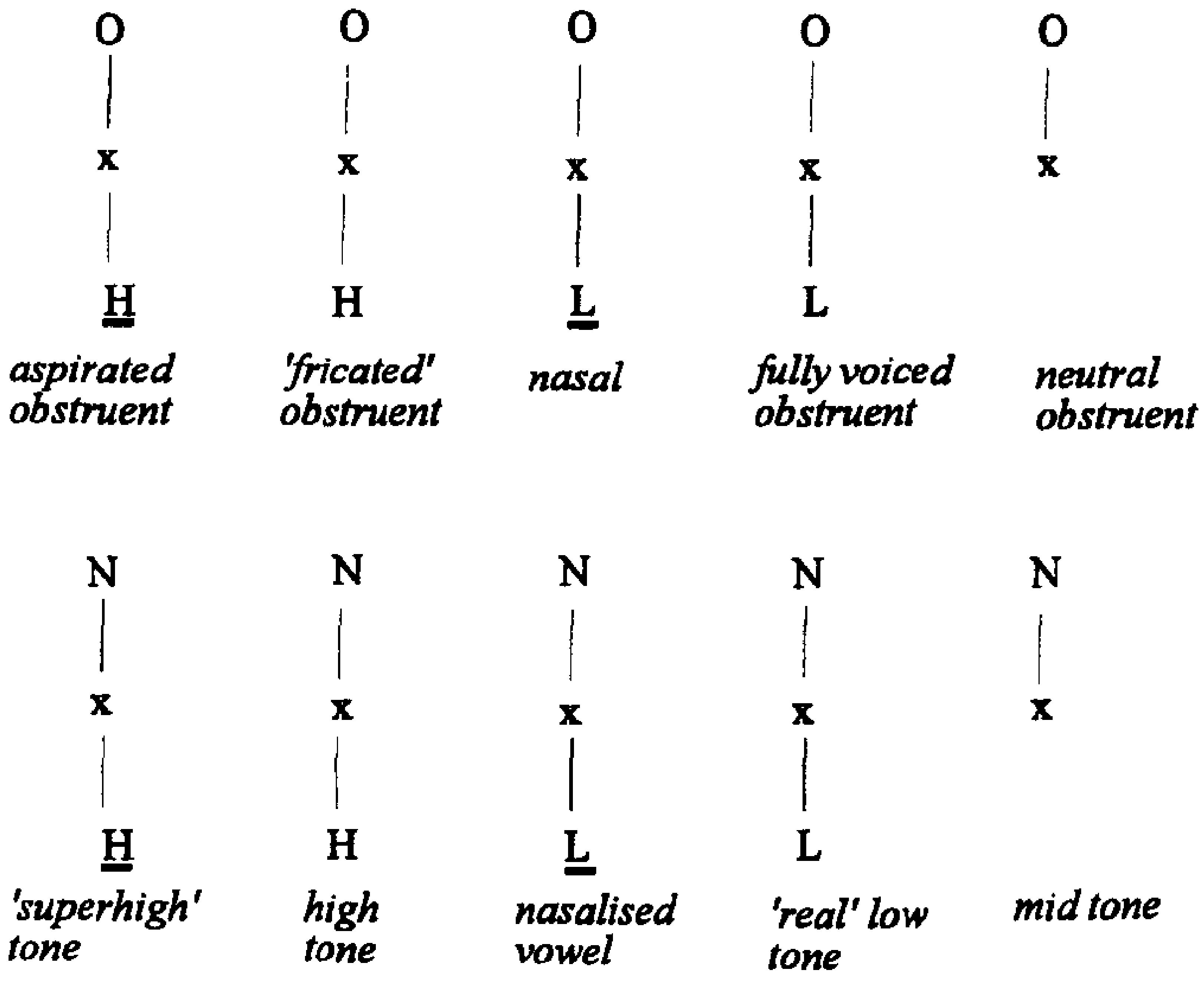
> ...a tone-bearing unit may have no tone associated with it underlyingly and perhaps even on the surface, in which case the TBU will be interpreted as having no instruction from the phonology. In this case it will be open to interpretation solely from the phonetics. (Hyman 1986: 115)

This statement refers to the 'widely accepted' generalisation that mid-tone is unmarked and unspecified. Within our present framework, we can certainly concur with the first part of this assertion: 'no tone' is synonymous with 'no laryngeal element'. But it is unclear how 'phonetics' could in any sense bridge a gap left by phonology, unless it constitutes a separate interpretive level, a idea which, as stated in §1.3.3 (p.27), seems to enrich the model unnecessarily from a psycholinguistic point of view. If, as argued in §5.4.2 (p.155), a prosodic domain may itself possess an acoustic signature, then a TBU lacking an active melodic object may nonetheless be interpreted because of its prosodic specification.

5.5.3 Logical conclusions about universality

In order to make the tonal asymmetries of Yorùbá and related languages explicit, we have adopted the idea that a prosodic domain may have an acoustic signature, independent of elemental content, and along the way we have accepted that [L] in onsets may be mapped either to a nasal or to a fully voiced obstruent, depending on its phonological headedness. Language-universally, these notions demand the set of logical possibilities for source elements in the several prosodic and isomeric guises delimited in (47). These are annotated with related acoustic properties. Some of these identities are more generally well-accepted than others.

(47)



A particular introduction needs to be made to some of the less familiar members of this family of representations. Firstly, the notion that headedness maps to phonetics is quite explicit here, following the remarks in §5.1.2 (pp.134) about the elemental presence of [L] in both nasals and fully voiced obstruents. In (47), there is a reverse isomeric polarity for [L] expressions from that proposed by Nasukawa (1998). This is not a crucial issue here, but the representations in (47) allow us to persist with the correlation of headedness and stricture that is implicit in the licensing constraints that apply to resonance elements (see §3.5, p.81). The reductionist endeavour has also been extended to the representations of [H] in onsets. An onset empty of source elements is genuinely neutral (and probably in a prosodically weak position: see Harris 1997). The prediction that the isomeric distinction of [H] and [H] obtains language-internally will have to await verification, but this contrast could now be used to distinguish between English-type aspirated obstruents and non-aspirated (under the working title 'fricated') obstruents, such as those of Tuscan Italian that possess an acoustic burst and are still capable of spirantisation under pressure, suggesting that they are not entirely neutral. This may be one step (but only one step) in the deposal of [h] from the elemental inventory. The logical possibility of unheaded [L] in a nucleus is not disowned universally, but if a 'real' low tone exists in a particular language, (tonogenesis suggests this may be the case for some east Asian languages, such as Vietnamese: see §2.4, p.59) then we expect phonological parallels with unheaded nuclear [H], parallels that we have seen do not exist in Yorùbá. Finally, the 'superhigh' tone is acknowledged to be speculative, and will have to await support from later analyses, presumably of languages bearing multiple tone heights.

5.6 The sound of prosody

Phonological evidence has now been mustered to show that high and low tones in Yorùbá must *not* be considered congeneric. The ability of high tone to show up in strong nuclear positions, together with the general robustness of this tone in west African languages, mark it out as having more melodic 'potential' than low. Low pitch, or falls, may signal prosodic, rather than melodic objects. The commonplace of using low pitch or falls signify 'The End' of prosodic sequences has previously been noted, so it may be possible to align this idea with considerations of morphosyntax and perception in a general linguistic context.

In some contexts, predictable interactions between phonology and morphosyntax are certainly discernable, notwithstanding our earlier remarks about bracketing paradoxes and prefixation. Morphosyntactic categorical boundaries are commonly cued by the lack of cross-boundary phonological operations, the most obvious example being the tendency for the edges of English intonational word-groups to align with syntactic edges.[35] Cross-boundary phonological juxtapositions generally show indifference to phonotactic constraints. But the *presence* of a phonological relationship may also indicate the limits of a morphosyntactic domain, even though this is a much less common event. One has to

---

[35] See also the reference to Neeleman & Weerman's (1997) work in §1.3.4, (p.31).

be careful when assigning this type of significance to Sandhi phenomena. Carr (1991: 48) has the following tempting example from Tyneside English:

(48)

(a) [ˈfɪʔ͡tɐ] *fitter* (b) [ˈfɪɹɐ] *fit her*

In this productive process, the 'Weakening' of the coronal stop is associated with the presence of a morphosyntactic boundary in (b); the lack of this process in (a) maps to a syntactic structure without this boundary.[36] This assertion is *not* Carr's focus in giving this example. His aim is rather to show that weakening is sensitive to the formation of a postlexical, rather than a lexical, foot. However, we might be induced at this point to argue that the net result for perception is the correlation of phonological activity to a morphosyntactic domain boundary. That is at least until we look further afield and see that first, this mapping only applies if the stop is coronal, and second, there exists the potential for a word-internal [ɹ] to be realised in a word like 'error'.[37]

Fortunately, a more convincing example exists closer to home, from the present west African perspective. Zsiga (1991) shows that Igbo vowel assimilation only takes place across a word boundary, rather than within a word. Assimilation is an exceptionless right-to-left process for non-high vowels across word boundaries. For high vowels in this position, it is optional. It never applies to intra-word vowels. Examples from Zsiga (op. cit.: 129) are given in (49).

(49)

Assimilation	[tʃɔ̀rɔ̀ égo]	→	[tʃɔ̀rèégó]	*want money*
	[ʊ́lɔ̀ à]	→	[ʊlàà]	*house this*

---

[36] The stop in (48(a)) is subject to Glottalisation, but this is an across-the-board phenomenon, and loses out to Weakening in environments which are identical save for the creation of a postlexical foot (Carr 1991: 49).

[37] Uvular and coronal rhotics may cohabit a Tyneside accent.

No assimilation	[tʃɔ̀-álá]	*link* + perfect
	[áwáɪ́]	*porridge*

The edge of a phonological domain, then, is *commonly* cued by an arbitrary juxtaposition of sounds, and this constitutes one regularity in the mapping between prosody and perception. But we can see that the mirror image of this relationship is occasionally attested. In Igbo, a word boundary is regularly the domain of vowel assimilation. Let us now examine the possibility that there may be a similar exoticism at the subword level in Yorùbá.

At higher levels of English prosodic structure, the edges of domains are usually cued by pitch change, the default being a fall. Where potential boundaries are cued neither by arbitrariness nor by pitch, then misunderstandings, and therefore jokes, become possible.

(50) *Scottish joke:*

Man in Glasgow shop:	'Is that a cake or a meringue?'
Shopkeeper:	'No, you're right, its a cake.'

It is understandable that the physiological design properties of the human respiratory system that make pitch declination the norm for utterances may generally result in the psychoacoustic mapping between a prosodic boundary and a drop in pitch. The typological exotic exemplar given by French, a possibly footless language in which the end of each word-group rises in pitch, does not undermine this generalisation. While the norm for a *subword* phonological juxtaposition of two syllables is essentially a relational affair - a foot - the proposition here is that one feature of Yorùbá's relatively exotic prosodic architecture is that it may use the low pitch cue which is normally associated with juncture at this subword level.

One further refinement of this statement is needed. The SW proposed here may not necessarily be marked by an 'edge-tone', or at least the pitch movement doesn't have to 'come' at the edge. In both the Yorùbá future/progressive verbal paradigms and the Ogori noun/adjective concatenations the temporal realisation of low pitch is not sited at the right edge of the domain.[38] This is not a convincing blanket refutation of the SW notion, as it is neither unique nor psycholinguistically odd. As Ladd (1996) points out, there are circumstances connected with *e-muet* in French where a particular emphatic contour, normally realised at phrase-edge, may become dislocated from that position. 'The last full vowel syllable has some special status as an anchor for tune-text association; we cannot simple say that the tones go to the edge of the phrase' (op. cit.: 57). Also, the general proposition that edge-tones may not be at the edge is not psychologically controversial: knowledge of language is unconscious and in the unconscious there is no time. Phonological word, using the term 'phonological' as defined in §1.3.4 above, like all prosodic domains, exists as an abstract entity bearing structural relationships to other abstract entities, a point which can often be obscured by the sequential models necessarily used to illustrate linguistic configurations. What we have here is a one-to-one mapping between a linear physical event (a drop in pitch) and a prosodic domain, which is psycholinguistically without temporal extension. There are limits to this point of principle because, as previously stated, phonology is grounded in sound (§1.3.4, p.32), but in all the present examples the acoustic event that cues the prosodic boundary does at least come *within* the prosodic constituent.

5.7 Summary

In this chapter, our attention has often been drawn to the tone hierarchy that obtains in Yorùbá and other African languages. It is now recapitulated one last time:

high >>> low >>> mid.

In §5.1, '>>>' was simply characterised as 'is more robust than', but a more specific explanation is now available. This hierarchy is never violated in African languages, as far as contemporary knowledge goes. The specific manifestations of it that have become variously relevant in the course of the present discussion are reviewed in (51), for the sake

---

[38] §5.4.3, example (35) (p.159) and §5.5.1, example (44) (p.165) respectively.

of clarity. All three specific languages are Benue-Congoid.

(51)

(a) To date, *all* African tone languages have been analysed as being possessed of either: (i) high tone only, or (ii) high and low tone, or even (iii) high, low and mid tone. No language has been analysed as possessing low tone only, or mid tone only, or low and mid tone.

(b) In Yorùbá, high tone does not show up in dependent nuclei (§5.4.1).

(c) In Yorùbá, high tone, but not low, signals a defective morpheme(§5.4.2).

(d) In Yorùbá, pronominals marked for high tone are invariant: low and mid toned pronominals may alternate (§5.4.3).

(e) In Ogori noun-adjective collocations, high tone survives at the expense of the other tones, while low tone survives at the expense of mid (§5.5.1).

(f) In Igede, high tone, but not low, spreads onto an object clitic (§5.5.2).

Most of these asymmetries can now be understood as related to the phonological underpinnings of tone heights, where high tone has elemental content ([H]), low tone signifies a prosodic constituent (SW) and mid tone remains precisely Ø. No account which retains congeneric [H] and [L] can do this automatically.

The hierarchy can now be read as the supremacy of the most highly specified, with elemental [H] imposing its signature on available positions, and SW showing up in morphological derivations where it is componential in one of the morphemes. This is not dissimilar to a comparison of the harmonic activity of nuclear resonance elements with the inertia of central vowels.

High tone, then, remains quintessentially elemental, being mapped to element [H], while low tone is said to reflect one facet of the idiosyncratic prosodic structure of the language. Prosodic idiosyncrasy is also evident in the way that Yorùbá displays constraints on word minima and maxima. The sub-prosodic-word SW sometimes operates in a 'fixing' role, to prevent impossible words, and sometimes has a lexical function. Again, this gives a phonological parallel to tonal asymmetries which has been previously unavailable.

Though there could be several problems awaiting the analyses of this chapter in a further study of Yorùbá prosody, we have at least begun to explain the tonal asymmetries that we set out to examine. The SW-hypothesis may require revision, but at this stage, a strong claim can be made that greater insights into the phonology of Yorùbá tone can only come from following the path chosen here and investigating the interactions of prosody and elements in the analysis of surface tones.

The endeavour in this chapter was originally inspired by the unexpected results of both the adult and infant perception tests detailed in chapter 4. If the analysis holds, we have a very simple reason for the lack of perception of any tone other than high tone out of a prosodic context: [H] is all there is.

Chapter 6

Contexts

6.1 Introduction

The preceding two chapters have served to exemplify an approach to acquisitional studies that freely combines typology, theory and psychophysical tests. There is of course no reason why such an approach need be limited to the acquisition of lexical tones and phonation. This chapter takes the opportunity to set what has gone before in a variety of contexts. One recurring theme of the previous chapters has been to emphasise that there is no guarantee that there is a one-hundred percent reliable mapping between phonological primes and their language-specific cues: the Yorùbá perception tests of chapter 4 and the correllated phonological analysis of chapter 5 concurred in the assertion that, for this particular language, low pitch does not cue element [L]. This kind of finding is consistent with moves to reduce the elemental inventory to those primes which are observably fully active ([A] [I] [U] [H] [L]: see §1.4, p.35), as it allows other extant phonological parameters (isomerism, prosodic position...) more of a say in the phonological structure of a specific language.

The immediately following section (§6.2) is a statement of a general psychological context for the phonological model adopted here, and this includes a psycholinguistic apologia for this reductionist drift. The proposal is then advanced (§6.3) that variability of assignment of a reduced elemental set during acquisition may account for some adult language-specific differences. This notion is examined in the light of coarticulations (§6.3.1), realisations that may be innocent of resonance elements (§6.3.2), and some original evidence from the transient consonant harmony stage which children variously exhibit (§6.3.3.). Elemental reduction is also briefly supported with a look at the various manifestations of rhoticity (§6.3.4). Section 6.4 underlines the need to be aware that non-phonological systems, both mechanical and systemic mentalistic though non-linguistic, may impinge on speech utterances, and that the acquisition of these is a separate issue from the acquisition of phonology. Section 6.5 is a note on the hitherto unmentioned correlation of perception and production in acquisition. The chronology of the acquisitional events that are relevant to the present discussion is laid out in §6.6, and this chapter is summarised in §6.7.

6.2. Elements and psychology

All perception is interpretive: objects must be represented in a suitable form to be available to a particular mode of computation. That there will be variance within universal boundaries for all interpretive routines is unsurprising: a psyche is of biological origin no less than a respiratory system is, and its reflective power reflects a shared planetary background as well as an individual's ontogeny.[1]

Symbolisation is one interpretive strategy available to human mentation. One of the most robust and useful tenets of element theory has been the identification of the tricorn vowel system as primal. The abstract patterns 'mAss, dIp' and 'rUmp' (Harris & Lindsey 1995: 53) would surely be assignable to the Tree of Life by any competent Qabalist. It is not frivolous to suggest that some notion of 'Trinity' may show up in both mysticism and linguistics: all that this suggests is that we use trifurcation for at least two mental routines. It may be easier to swallow the idea in concert with another repeated theme in our perceived existence. Binarity is clearly rooted in our biology, and while its most far reaching practical application is of course currently to be found in computer technology, it is also crucial to a great many of our widely accepted linguistic principles. Within phonology, there is a ubiquity of maximally binary structure in 'syllables' and, for some phonologists, we are getting close to a principled account wherein the only constituents are an endlessly repeated onset/nucleus pairing (Lowenstamm 1996). At the segmental level, binarity has been exploited to the utmost within 'Radical CV Phonology' wherein all 'exponents of phonological categories', as they are termed in this system, are ultimately

---

[1] See also the remarks in §1.3.4 (pp.30-31) about evolution and abstractness in phonology.

derived from combinations of two abstract primes, C and V (van der Hulst 1989, 1994,1996). These primes derive, with the inclusion of headedness relationships, an 'almost complete model of segmental structure' (1994: 473), but interestingly from the point of view of this present discussion recent developments of this model have also taken into account the iterated tripartite syntax familiar from sentential structure which utilises 'specifier', 'head' and 'complement' nodes (1996). For phonological representations, even from models which are to some extent agonistic, there is a counterpoint of duple and triple oppositions.

Our contention here is that it is not surprising to find that our grammatical model includes basic structural aspects that are also to be found in other modes of human mentation. It is orthodox that language is genetically endowed and attuned within boundaries set by Universal Grammar (Chomsky 1986, 1988, etc., etc.), and so we are able to predict that grammatical constructs will be rooted in the abstractions which our psyches have already, in the course of evolution, generated from their interaction with the universe. Recalling that our phonology is modular, and that we have maintained again and again that it can best be regarded as minimally bi-modular, with particular reference to acquisition, (§2.3.2, pp.50-54), then we expect that the ability to recognise the patterns present in the module(s) will be acquired, in a biologically determined fashion, by the interaction of endowment with both common acquisitional experience and the linguistic environment.

The degree of abstraction that is achieved by this interplay is important. It allows us to contend that, while there will be common 'paths' of acquisition cross-linguistically, these can be allowed to diverge at various points.

A *caveat* should be aired, though, about the nature of the abstractions that may be utilised in any particular module. Peters (1973) delivers a logically plausible account of downdrift in lexical tone languages wherein numerical values are assigned to H and L tones, and arithmetical computations deliver the surface shape of the utterance. Odden (1995: 454), following Clements, points out that giving phonologies the power to perform integer arithmetic is undesirable, as this is 'a level of descriptive power which is unwarranted'.

In contrast with the disjunction of a putative mental arithmetic facility for downdrift within the phonological component, it is positively encouraging for our present model to find a juxtaposition between two and three term oppositions as a repeated *leitmotif* not only in competing phonological accounts, but also in contemporary theories of syntax. In element theory, the obvious bearers of duality and trinity are respectively [H L] and [A I U], and it is precisely these five elements which have been seen to be the main actors on the phonological stage (see §1.4, p.35).

6.3. Potential for variation in the assignment of elements during acquisition

6.3.0 Following on from the proposal that there is an innate psychic commonality to be taken into consideration when considering the psycholinguistic advent of phonology, it is worth considering that there may be early acquisitional experiences that may be related both to the 'natural' assignation of elements to phonetic signatures, and also to some departures from the norm in this respect.

The concept behind the patterns which are at the heart of element theory and their potential correlation with acoustic phonetics is not, of course, new. The Prague School linguists were using melodic features defined in acoustic terms to distinguish phonological oppositions, and though for some time this approach ran into general disfavour, its resurrection is now thoroughly under way, at least on the east of the Atlantic Ocean. Dependency Phonology explicitly relates its own combinatory primes to Jakobsonian correlates. Anderson & Ewen (1987: 234) reintroduce in principle the feature [grave] for sounds in which the lower part of the spectrum predominates, in order to account for the recurrence of labials and velars as a natural class: this is then identified with '...the (DP)[2] component |u|...also acoustically glossed as 'gravity'...'. Linguists working in Government Phonology also acknowledge their pedigree. In using acoustics, rather than articulation or auditory perception, as his first base for an elemental gloss, Harris (1994: 107) cites a particularly illustrious source: '...this stance was explicitly adopted by Roman Jakobson

---

[2] Parenthesis added.

in early distinctive-feature theory'. More recent GP work includes a discussion of coronality that takes its 'main intuition' from the same source and begins with the grouping of '...alveolars (/t/), palatals (/c/) and /i/ as acute' (Foltin 1996: 63).

But such unity of inspiration is belied by a great diversity of detailed reinterpretation. Without spreading the theoretical net any wider than a discussion of element theory in GP, it is soon possible to encounter some fairly basic areas where there is no consensus, the representation of coronals being a case in point. Analytical diversity is also obvious in the analysis of element [U], which is in orthodoxy said to map one-to-one with labial(-velar)s, but as discussed in the immediately following section, has been analysed as componential in velars or uvulars in at least one language. In both these cases, it may be possible to cite language-specific acquisitional experience as an explanation for variable assignation.

6.3.1 Phonological acquisition and food

To first illustrate this point, we can further extrapolate from some facts already discussed by Jakobson (1962) and extend his account to include the route to phonological maturity that is predicted by the element-theoretic model. A suckling infant will often accompany this activity with a nasal murmur: both acoustically and in an articulatory description, this bears the characteristics of a velar nasal. If removal of the nipple gives cause for complaint, this will be associated with a labial closure. Jakobson goes on to propose that this vocalisation, in concert with the primal CV syllable, leads infants in the later course of their development (at 1:0:0 to 1:3:0) to produce the utterance [mama] as an indication that they feel it is about time they were given something to eat.[3] So, Jakobson proposes, we have an account of the ubiquity of the association of this phonetic shape with the (shortcomings of) the provider of sustenance. The present focus, of course, is not the acquisition of meanings but of pattern-recognition, and we have proposed that this is under way at six months of age. If this same feeding experience is cited as formative for

---

[3] A rider may be added to the effect that the common minimality condition that a word must be a (binuclear) foot also has a role here.

phonological acquisition then we have the beginnings of an explanation for the typological unmarkedness of the labial-velar coarticulation and the related simplex nature of its usual phonological correlate. The close association, during feeding, of the psychoacoustic perception of the signatures of both labial and velar closures, regardless of native language, may be an instance of linguistic knowledge developing from common experience, with putative language-specific and language-driven refinements taking place at a later stage to allow the phonological object mapped to these closures to optionally attach to one or the other.

Back high rounded vowels, utilising both constrictions, (as opposed to back high unrounded vowels, which only use one) are very much the 'unmarked' realisations of element [U] in nuclear position. The same goes for glides. A glance at Maddieson's (1984)[4] list of 317 languages reveals that the doubly-constricted [w] type glide appears 233 times, with [ɥ] (front rounded) type glides and [ɰ] (back unrounded) type glides taken together showing up no more than 10 times (six for the first and four for the second). A story which includes common acquisitional experience allows us to see why an apparent phonetic complexity (a double articulation) may line up straightforwardly with a phonological singularity.

Independently motivated research loosens any overly simplistic ties between [U] and 'round'. Harris & Lindsey (1993), arguing against the existence of a systematic level of phonetic interpretation, provide us with the following statement:

> 'In the articulatory implementation of...[U]..., labial activity is but one of the factors that contribute to the overall size of the oral cavity. One implication of this is that lip-rounding is not an articulatory prerequisite in the production of vocalic expressions containing [U].' (Harris & Lindsey (1993: 366)

Though lip postulation is only one of the articulatory factors associated with [U], this

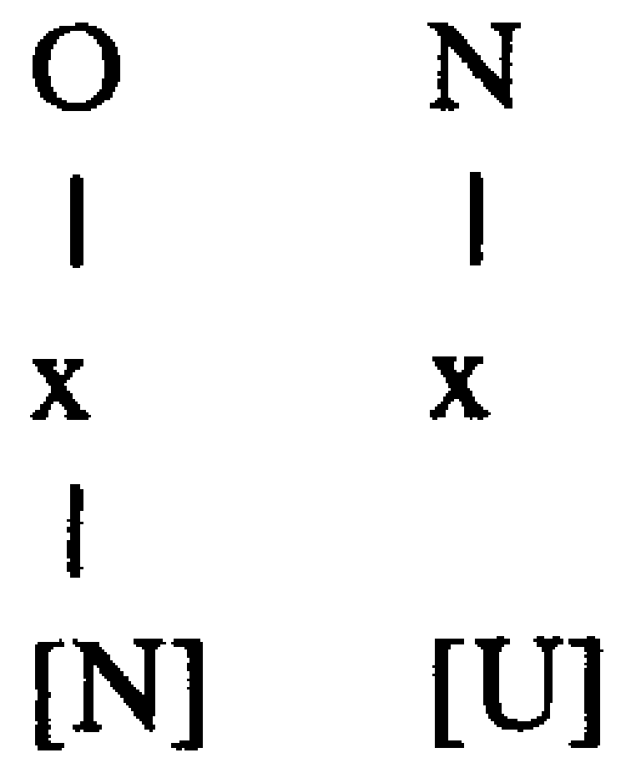
---

[4] Though this work of reference is couched in exclusively phonemic terms, this does not bar us from drawing from it an overwhelming statistic of this type.

account would seem to allow at least limited variability in the assignation of melodic objects to phonetic signatures by speakers of different languages. Let us extend this proposal to the point where [U] may logically be realised in language *p* (such as English) as a labial-velar, and language *q* as, say, a velar. Acknowledging this variation would allow the unproblematic incorporation of a language such as Japanese, with its (for most speakers) high back unrounded vowel, into the family of phonological /a e i o u/ languages. Like roughly one-quarter of the other languages of the world, it simply uses nuclear [A], [I], [U] with colour tier fusion.

GP analyses exist which present independent facts explicitly identifying Japanese as a *q*-type (non-round [U]) language. The Japanese syllabic nasal inherits its place specification from the content of a following onset, but will word-finally have a back quality (velar or uvular). Yoshida (1995) analyses this nasal as having the underlying structure given in (1):

(1)



Supporting this analysis, Cobb (1995) cites evidence[5] that the Japanese negative affix was historically (in Old Standard Yamato) [nu], and that this same affix currently bears the reflex [N] in both western Japanese dialects and Modern Standard Japanese constructs such as /arimase-N/ ('there is' + *neg.*). The cohabitation of nasality and [U] lends a back, rather than a labial, quality to the nasal 'mora'. Cobb also cites as corroborative evidence of the existence of [U] in /N/ a statistical gap in the occurrence of the Kana form [nɯ]. Of 10,957 instances of the five nasal-initial Kana sequences [na], [ni], [nɯ], [ne] and [no] in a string of Japanese, only 298 (2.7%) occurences of [nɯ] show up: the range of percentages for the other four forms is 13% - 30%.

---

[5] For references see Cobb (1995: 33).

The important point is that we can, if we wish, accept the analysis delivered by Yoshida, Cobb and the sources they cite while maintaining that this analysis holds only analogical implications for the analysis of any other language: such analogies have no principled reason to hold. The identification of [U] in Japanese [ŋ] / [ɴ] (and presumably in [ɯ]) may be supportable but this has no implications for, say, English [ŋ]. English velars have been most frequently analysed as reflexes of some uniquely underspecified object such as the 'background' element [@] (Harris & Lindsey 1995), or as being possessed of 'empty-headedness' (Ritter 1996). These competing accounts will be reviewed with regard to phonological complexity in the following section. Meanwhile, if Japanese and English babies alike utilise a phonologically simplex [U] but depart in their assignation of this element to a phonetic signature, this constitutes support for the need to look at such assignations on a language-by-language basis.

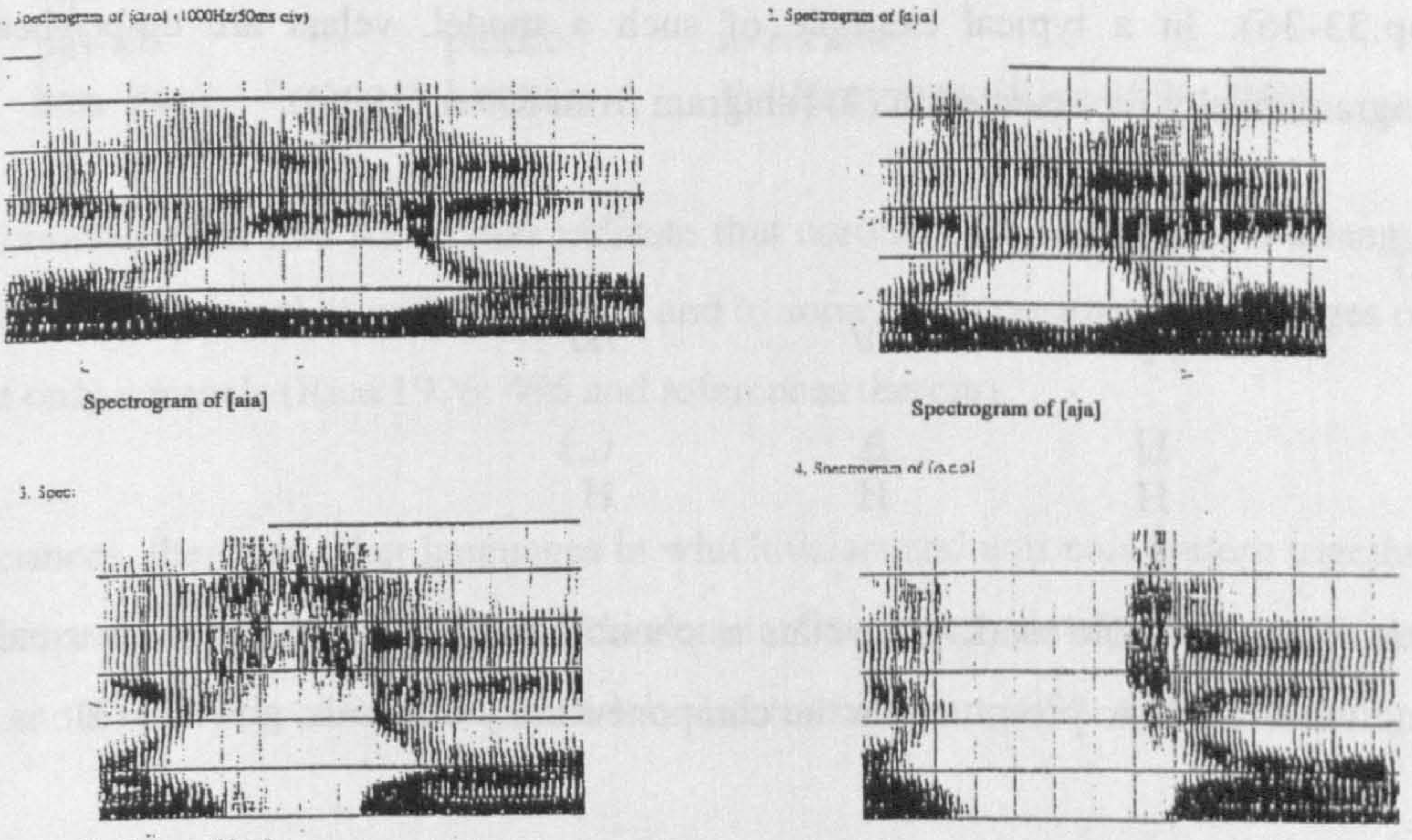
This story of variable phonetic implementation does not undermine any of the relational regularities that do exist, such as the correlation of [A] with non-high vowels, or of course of nuclear [U] with high back rounded vowels. There will indeed be predictable regularities in the acoustic/articulatory correlates of elements, but there will also be a degree of language-specific variation, the origin of which should be traceable in acquisition. Each language brooks a separate research enterprise to ascertain its degree of divergence from some 'norm' of element mapping given by UG.

6.3.2 Variable placelessness

To reiterate once more, in orthodox element theory not all elements are created equal. [@], in particular, has been identified as a 'background element', a 'blank canvas'. It is only a short step from here to the idea that [@] is not an element at all, so is more clearly orthographically represented as [ ]; truly featureless. After briefly reviewing these two options, this section will go on to propose that phonological featurelessness can be severally phonetically assigned, just like a strongly identifiable object such as [U]. If this variability of assignation is accepted, it should also have a similarly acquisitional origin.

Evidence for lack of specificity generally includes tendencies to assimilate, lenite or elide under pressure. All these types of evidence are grouped together in this section.

The phonetic correlate given in Harris (1994) that supports [@] as componential in (English) velars represents a departure from the norm. Usually in this analysis a certain amount of acoustic invariance is invoked, at least as corroborative evidence, to support the occurrence of identical elements in different prosodic positions. The presence of [I] in both palatal vowels and palatal glides is a case in point. While elements remain abstract, there can be no mistaking the connection between F1 and F2 values and the patterns underlying [A], [I] and [U]. The four spectrograms in Figure 6.1 show a neat correspondence between F1 and F2 values not only for palatal vocoids but also for the locus (target) frequencies for palatal closures.[6]



Spectrogram of [aia]

Spectrogram of [aja]

Spectrogram of [aça]

Spectrogram of [aca]

Figure 6.1 Spectrograms of four palatal articulations: vowel, glide, fricative and stop.

An appeal to acoustics is no help, however, with the identification of [@] in velars. The

---

[6] Spectrograms courtesy of Mark Huckvale, UCL.

acoustic correlate relevant to [@] in nuclei is said to be the 'neutral tube' formant pattern typical of central vowels, i.e. F1 at around 500Hz and higher formants spaced out roughly evenly at 1 kHz intervals (Choo & Huckvale 1993, Harris 1994: 140). Nothing of this kind can be found in the track of a velar consonant: F2 in this closure is aiming at a locus frequency of +3kHz. It is true that acoustic signatures were never intended to be pivotal in the identification of elements, but they are certainly cited as circumstantial evidence when available, and this evidence is, in the case of velars, non-existent. So this time the story is firmly articulatory and needs to appeal to a process of elimination: 'Since [@] independently defines a sound which is articulated as dorsal (produced with the tongue body), non-palatal, non-labial and non-open, it can be taken to specify velarity in consonants' (Harris 1994: 119).

As an alternative to this approach, the most radical reductionist literature within element theory has espoused the five-element inventory which was supported as irreducible in §1.4 (pp.33-36). In a typical example of such a model, velars are empty-headed, as diagramatically represented in (2) (diagram from Ritter (1996)):[7]

(2)

/p/	/t/	/k/
$\underline{U}$	$\underline{A}$	(_)
H	H	H

This system therefore marks out velars as phonologically exceptional (as is usual in GP), a notion even more prominent in the componentiality of nasals, given in (3):

(3)

/m/	/n/	/ŋ/
$\underline{L}$	$\underline{L}$	$\underline{L}$
U	A	

---

[7] Ritter herself has since amended this analysis to include (for dorsals) element [v] (Ritter 1997).

While it would be neat to succeed in assigning elements to universal acoustic/articulatory correlates in this way, a look at individual language facts unfortunately makes this endeavour look overly idealistic. If we accept that a tendency to assimilate is evidence of non-specificity, many sources present evidence that identifies coronals, rather than velars, as being relatively malleable under pressure. Phonologically, 'coronal' displays all kinds of unique behavioural traits, which are made explicit in Paradis and Prunet (1991). These include its high cross-linguistic incidence, its susceptibility to assimilation, and its transparency to harmonic processes. Place assimilation of coronals in English is familiar ([tem bɜːdz], [ten dʌks], [teŋ giːs]),[8] but this phenomenon is not confined to Germanic languages. Rice (1996), following Cho (1988), has the following examples from Korean.

(4)

kotpalo	→	koppalo	*straight*	
han-ben	→	hamben	*once*	
pat-ko	→	pakko	*to receive*	
han-kaŋ	→	haŋkaŋ	*the Han river*	(Rice 1996: 494)

Distributional facts also sometimes indicate that coronals are unmarked: if a language exhibits only one nasal, it will be coronal, and in some weak positions[9], languages often permit only coronals (Rice 1996: 496 and references therein).

Furthermore, there are other languages in which velars and coronals pattern together in these respects. In Chaha, a Southern Ethiosemitic language, palatalisation can generally dock onto a velar or a coronal but not a labial (Lowenstamm 1996: 424), indicating that

---

[8] It is proposed in §6.4 (pp. 194-196) that these assimilations may not be due to elemental specificity at all, but for now the evidence is reported in the spirit in which it is usually offered.

[9] 'Some weak positions' is deliberately unspecific: the facts here are taken from Rice (1996) who characterises these positions as 'dominated by rhyme nodes': they are in fact final cononants, so in our terms they are onsets licensed by final empty nuclei (at the end of a licensing path which begins in the foot-initial nucleus).

in this language labials may be more highly specified than coronals *or* velars.[10]

Because evidence exists that marks out both velars and coronals as lacking specificity, Rice (1996) assumes this to be a contradiction in need of explanation. She hypothesises that coronals surface by dint of a 'fill-in of the unmarked feature', whereas velars are the result of a 'failure of fill-in of the unmarked feature, with [phonetic] interpretation of the node' (Rice 1996: 497). The proposal that there exists an independent level of phonetic interpretation would need a lot of corroboration in view of the undesirable enrichment of the model this involves, and is certainly contrary to the spirit of the present argument. In any case, if it is accepted that assimilability is linked to non-specificity, it is also possible to find examples of languages where labials assimilate, a fact which undermines the basic assumption of Rice's argument. A fragment of a Japanese verbal inflection paradigm is presented in (5).[11]

(5)

kam	+	ta	→	kanda	*chewed*
sin	+	ta	→	sinda	*died*
tob	+	ta	→	tonda	*flew*

If we leave assimilation to one side, and look instead for a tendency to lenite as a sign of non-specificity, we see that the weakness of coronals in Germanic languages and Korean is paralleled not only by assimilation of velars (in, say, Chaha) but also by the elision of velars in a language like Turkish, where under suffixation [inek] + [-i] is realised as [ineØi] (*cow*). Labials, while cross-linguistically statistically resistant to assimilation, are nevertheless subject to parallel processes of lenition in diachronic shifts. Two of these are exemplified in Lass (1984: 336-7) and reproduced below as (6). The second language in both (a) and (b) betrays ossification at two different stages in the same lenition trajectory from starting point [p] (serial spirantisation/debuccalisation/elision).

---

[10] Another observation indicates that any account of this process which does not consider prosody is too simplistic, as palatalisation in Chaha is subject to the further condition that it cannot take place onto root-initial coronals (Lowenstamm 1996: 427).

[11] Data from Nasukawa (1997). His work was discussed in §5.2.2, (p.134).

(6)

(a) *Dravidian:*

Tamil	Kannada	*gloss*
pɔku	hɔgu	*go*
pɑl	hɑlu	*milk*
pal	halu	*tooth*

(b) *Indo-European:*

Latin	Old Irish	*gloss*
porkus	ore	*pig*
pater	aθir	*father*
piscis	iask	*fish*

The important point here is the lack of clarity from cross-linguistic evidence as to what constitutes a 'weak' stop. It is therefore arguable that non-specificity, with its attendant lack of elemental complexity, may be a characteristic of diverse phonetic reflexes. Extrapolating language-specific evidence as indicative of universals may be a mistake. It is also generally an unreliable and unrealistically phonemicist position to identify melodic constructs out of context.

In §6.3.1 (p.182), we were faced with the conclusion that while nuclear [U] would typically map to round high back vowels, at least one other option existed. Nuclear [A] appeared to be more predictable. The question we are now required to ask once more is exactly why elemental identity is more transparent in some cases than others; this section opened by noting the exemplary behavior of [I] with regard to its acoustic signature, and this has been contrasted with the less satisfactory stories that attach to [@] (or [ ]). One possible answer to this problem could be the acquisitional mapping between grammatical objects of equal status and psychoacoustic triggers of varying degrees of salience. This predicts both commonality and specificity, and allows element theory to possess, at least temporarily, what mathematicians call 'areas of good agreement'. This investigation cannot then be furthered without again undertaking research into a genealogically large number of languages, to determine just where each disparity lies. A language-universal account is going to have to acknowledge all potential diversifications.[12]

In this regard, it is also worth restating that we may sometimes, in the steady-state language, need to disentangle an element from its prosodic web to see if it really is what we think it is. If the empirical reasons given here for the possibility of variable phonetic interpretation are accepted, then from a theoretical point of view, this contention also appears to receive accord in places.

Brockhaus, Ingleby & Chalfont (1996), for instance, while maintaining the singularity of elements, postulate four conditions in which an element must potentially be identified:

(i) as a head in nuclear positions
(ii) as a dependent in nuclear positions
(iii) as a head in non-nuclear positions
(iv) as a dependent in non-nuclear positions.

The proposal that the interaction of prosody and melody can affect phonetic expression, just as isomeric or licensing relationships can, proved useful in accounting for phonological distributions and for phonological anomalies in the previous chapter. There is nothing anywhere in this theoretical classification that directly weighs against the idea that prosody *without* melody may have an acoustic signature.

6.3.3 Placelessness and consonant harmony

Some acquisitional evidence is now presented that bears on the issue of the relative weakness of coronals and velars, and the relevance of prosodic positions. Despite the well-known tendency of coronal (stop)s to cede place specificity in adult English

---

[12] There is also often a weakness in the way in which 'phonetic expressions' are themselves modelled in representations. A good deal of the time they look a lot like phonemes, and take no account of either prosodic position or the idiosyncratic phonological requirements of a specific language, dialect, or idiolect. This is partly due to the confusing use of the word 'phonetic' that was identifed in §1.3.3 (pp.23-27). No useful theory can evolve from simply taking the 'phonemes' of a language and translating them into elemental terms in this way.

assimilations, young acquirers regularly go through a stage when they assimilate velars. Two classic patterns are well-documented, neither of which exhibit directional constraints:

(i) Velars harmonise with coronals or labials. Classic examples are [pæp] 'cap', [tæt] 'cat', [dɒd] 'dog', [mʌb] 'mug'. 'Kick' is still [kɪk], showing clearly that velars are realisable and that this is a harmonic process.

(ii) Velars become coronal if the consonants within the domain are not exclusively velar. The result is exemplified by data identical to that in (i) above, save that 'cap' is now [tæp] and 'mug' is now [mʌd].[13]

In textbook cases, these conditions are said to obtain within words without exceptions, but life is not always this simple. The word list in (7) (over) presents forms produced by a subject aged 1:11:12 and betrays a more complex system.[14] All of these words (except 'cop') were uttered while the subject was looking at a picture book: this sidesteps the possibility (inherent if the child were to be asked to repeat words) that imitative skills could render the data experimentally noisy.

This inventory of utterances can be interpreted as follows. (a) shows simply that velars are part of the repertoire, (b) that neither labials nor coronals assimilate within a CVC word and (c) that, also within this structure, velars do not assimilate to labials. By contrast, strong evidence of harmony is present in (d): in these words, velars harmonise, regardless of directionality, with coronals. So far things are as simple as they have been portrayed in the 'classic' cases. However, a consideration of the data in (e) and (f) reveals a different

---

[13] All examples in (i) and (ii) here are taken from data presented as analytical exercises to first-year speech science undergraduates at UCL.

[14] Thanks to Robin Barker and his parents for providing these examples. Transcriptions of this type are always open to criticism, but those presented in (7) seem close enough to make the necessary points clear. ['ɗĩːkĩŋ] 'drinking' perhaps needs particular support. This is intended to represent the fact that there is a velar quality to the second syllable that is lacking in the first, which becomes relevant to our discussion of parametrisation.

(7)

(a)	kɪk	'kick'
(b)	tɒp	'top'
	pɒt	'pot'
(c)	pɪg	'pig'
	pɒg	'frog'
	kɒp	'cop'
(d)	dɒd	'dog'
	dʌd	'jug'
	tɒt	'cot'
	dʊəs	'goose'
(e)	'dɪ̃:kɪ̃ŋ	'drinking'
	'dɒgi	'doggie'
(f)	'bʌkɪk	'bucket'
	'tɒkək	'chocolate'
	'ti:kɒk	'tick-tock'

pattern. (e) shows that velars *can* cohabit 'word' with coronals: the utterances [dɒd] and ['dɒgi], produced within half a second or so of each other, are particularly convincing evidence. Moreover, the examples in (f) show assimilation not of velars, but of coronals. The vital ingredient which apparently shifts the pattern is the presence of the second filled nucleus, giving a 'full' trochee. Harmonic processes are prosodically conditioned for this child.

The facts can be summarised as follows:

(i) In words having one full and one 'dull' syllable, velars harmonise with coronals.
(ii) If there are two full syllables, the harmonic process is absent.
(iii) In words having two full and one 'dull' syllable, coronals harmonise with velars.

CVCØ is always the domain for consonant harmony, but the melodic target of the harmonic process varies depending on the overall word structure.

The data as a whole provides firm evidence of prosodic parameter setting. The final empty nucleus parameter is ON, while 'branching onset' is as yet [OFF]: witness [pɒg] ('frog') and [dɪːkɪŋ] ('drinking'). 'Branching rhyme' is also [OFF], the child incorporating the stray skeletal point of the rhymal adjunct in [ˈdĩːkĩŋ] into a branching nucleus which exhibits the nasality that will ultimately be present in the adjunct. Here, the first syllable of the child's pronunciation is devoid of phonetic velarity, unlike the second. In the adult language, velarity is not directly a-licensed in the coda of the first syllable but via the following onset which *does* license velarity: no coda, no velarity.[15]

This much is arguably derivable from the evidence in (7). We now unfurl a big banner labelled 'Speculation', and under it, propose the following: if 'velar' is the factory setting for non-specificity of place, the data in (7) betray a developmental stage where the acquirer is somewhere between this initial state and the steady state for English, when 'coronal' will be non-specific.[16]

It is well-known that despite the wide attestation of harmony between obstruents at the time of single-word utterances, evidence of this phenomenon in adult languages is scarce.[17] By contrast, vowel harmony systems are relatively widely distributed. Since harmonic relations are contracted at a projected level of the prosodic hierarchy, and since only nuclei project, the adult facts are readily understandable, the rarer consonant

---

[15] Interestingly, no further (rhymal) parameter setting by the child would be needed for him to qualify as a speaker of Malay, Yucatec Maya (Mexico), or Luo (Eastern Sudanic). This is a further small confirmation of the veracity of the arguments put forward in Harris (1994: 160ff.).

[16] A unified account of the data demands that [ˈtɪːkɒk] ('tick-tock') is analysed just like [ˈbʌkɪk] ('bucket'), and [ˈtɒkək] ('chocolate'). An objection to this putative syllabification by the child can be found in the general proscription against [ɒ] in the subordinate nuclear position in English. However, the existence of words like 'despot' and 'Ascot' means that either this proscription has exceptions, or that these words are analytic. A further possibility is that our subject has yet to regularise his English weak-vowel system.

[17] One example: Chumash (Amerindian) has a domain in which either all, or no, sibilants are palatal (Poser 1982).

harmony systems needing to exploit more complex licensing paths via intervening nuclei. Because child consonant harmony peters out at around the age that phrasal utterances are coming in (at around 2:0:0), it is possible that harmonic relations between onsets across intervening nuclei in the immature language generally cease to be contracted as phonological phrase level projections become available.

6.3.4 Varieties of /r/

Further support for the potential variation of the phonetic manifestations of nonspecific place can be obtained from a brief review of the world's rhotics. The only thing that these have in common as a phonetic 'class' is that they have very little in common. They are therefore potentially useful evidence for the evaluation of the claim that some phonetic objects may be elementally unspecified for resonance.

The diversity of phonetic rhotics is clearly laid out in Ladefoged & Maddieson (1996): after a detailed description of their manifestations, which include approximants, trills, taps and fricatives ranging from dental to uvular in place of articulation, the conclusion is drawn that 'the overall unity of the group seems to rest mostly on the historical connections between ... subgroups, and on the choice of the letter 'r' to represent them all' (Ladefoged & Maddieson (1996: 245).[18]

As far as phonology goes, though, rhotics do have an overall identity. To cite one example, in Europe there are a cluster of dialects of French, German, English and southern Swedish sited within a clear isoglottic boundary that possess overwhelmingly uvular rhotic realisations. This is reflected in the acoustic signals of these phones, which lack the low third formant traditionally associated with /r/. The phonological distribution of uvular rhotics in these languages is identical to that of coronal rhotics in their dialectal cousins. They share all phonotactic constraints, and may even alternate with coronals within a single idiolect, subject to regional factors. They may be in free variation, or in

---

[18] For a proposal about the melodic makeup of three of the 'manner' variations of /r/ which in fact uses resonance elements, see Harris (1994: 259).

complementary distribution: in some Dutch, /r/ is uvular prevocalically, and coronal (retroflex) before a consonant or boundary. If we credit any predictable mapping at all between psychoacoustics and phonology, this kind of diversity certainly argues for lack of place-specificity. Absolving rhotics from a commitment to a particular resonance element in these cases gives less of a problem accounting for these phonological parallels.

What of languages which utilise contrastive rhotics? Around 13% of languages do so (Maddieson 1984) and there is an overwhelming tendency to contrast parameters other than place, once again underpinning the notion of place-non-specificity in the coronal realisations of 'r-sounds' familiar to most English speakers. In Spanish, for instance, one intra-rhotic contrast is possible. An intervocalic tap may contrast with a trill, resulting in minimal pairs such as [ˈpeɾa] 'pear' and [ˈpera] 'bitch'. In all other positions, the trill and tap are in complementary distribution, the trill showing in onset-heads ([ˈroxo] 'red', [ˈrweɣo] 'I beg') and the tap in weaker prosodic positions - onset-dependents ([ˈtɾutʃa] 'trout', rhymal adjuncts ([ˈaɾte] 'art') and onsets licensed by final empty nuclei ([daɾ] 'give').[19]

It may not even be necessary to look to the melodic component to account for this contrast if it is argued that the trill is a geminate tap. However, it must then be acknowledged that Spanish generally lacks the geminates that ideally would support this proposal. A comparison with Italian makes this clear. Italian shares a close genealogy with Spanish, and in many cases uses a more-or-less identical vocabulary. However, in environments where Italian will freely exploit contrastive length in consonants, Spanish must draw on a quite different etymological stock, as illustrated by the data in (8).

(8)

Italian	*Spanish*	*Gloss*
ˈgrato	agradeˈθido	'grateful'
ˈgratto	rasˈkaro	'I scratch'

---

[19] A discussion of onset-licensing is well outside the scope of this chapter, but will be found, among other places, in Kaye (1990).

'tono	'tono	'tone'
'tonno	a'tun	'tuna'
'pene	'pene	'penis'
'penne	'plumas	'feathers '

Though analysing the Spanish contrast as one of prosodic quantity therefore marks the trill out as the single geminate in the language, there is plenty of evidence that this is not phonologically bizarre. Firstly, other languages which unequivocally *do* have geminates such as Afar (Afro-Asiatic) and Shilluk (Nilo-Saharan) exploit precisely the same phonetic contrast - the trill for the geminate and the tap for the single position. Secondly, the trill is historically derived from a sequence of [r] + coronal stop (Ladefoged & Maddieson 1996: 237). Thirdly, languages are often constrained in the melodic content of geminates: in Sesotho, for example, only sonorants geminate (Harris, pc).

In spite of the non-specificity of Europe's rhotics, it is of course possible for some 'r'-sounds to participate in a genuine place contrast. In Hausa, for instance, an alveolar tap (or short trill) contrasts with a post-alveolar tap (or approximant), and both may geminate (Ladefoged & Maddieson (1996: 237). Even so, the inventory can easily be accounted for by the inclusion of [I] in the makeup of the post-alveolars, without the necessity of proposing a separate resonance specification for the alveolars.

6.4 Mechanics and mimics

Elements seem to be uniformly at their best as an explanatory tool in accounting for long-distance relationships like vowel and nasal harmony and tone spreading. They also deal economically with the decomposition of vowels. In addition, it is sometimes possible to deliver a neat account of the participation of elements in strictly local processes. But in this final class of events, a transparent explanation is not always so reliably forthcoming.

Processes concerning consonant place of articulation, in particular, involve a less than obvious assignation of elements in some cases. On the plus side, Harris (1994: 118-119) gives an explication of English palatalisation using [I] spreading, and of historical vowel rounding (e.g. [wænt → wɒnt]) involving the spread of [U] from onset position and the consequent delinking of [I], an effect already predicted by the cohabitation of [I] and [U] on the same autosegmental tier. However, in our consideration of non-specificity of place in §6.3.2 (above) it became clear that there was a mismatch between the predicted universal 'weakness' of velars (whether or not they are empty- or [@]-headed) and the language-specific weakness of (say) English coronals. The proposal was offered that the variable assignation of elements to phonetic cues during acquisition could account for this. The acknowledgement that different languages may have different places of articulation that are melodically impoverished means that this particular aspect of element theory falls somewhat short of the explanatory power of, for instance, the tricorn vowel system.

It may be prudent, therefore, once again not to maintain a doctrinaire insistence that all 'segmental' contrasts are necessarily to be accounted for using elemental activity. A classic place assimilation paradigm for the nasal in the English 'RP' word 'sun' is presented by the compound nouns in (9).

(9)

(a) sʌm bed
(b) sʌɱ fæktə
(c) sʌn̪ θerəpi
(d) sʌn tæn
(e) sʌŋ kriːm

It seems possible that these assimilations of a non-specific nasal are not best accounted for by invoking elemental distinctions. An automatic, mechanical coarticulation seems like a simpler explanation, with the nasal remaining unspecified in all realisations. Browman & Goldstein (1989) use exactly these circumstances to support their notion of 'gestural' phonological primes. They have succeeded in making articulatory measurements showing that the 'closed alveolar' gesture associated with English [n] is still traceable in a running speech production of 'seven plus' where the most accurate transcription may be [sevɪm̩ plʌs]. The assimilated pronunciation is the result of masking of the closed alveolar gesture by the temporally adjacent labial fricative and stop. They cite a similar explanation for an assimilated pronunciation of [n] in'ten themes'. (1989: 215-219). This articulatory evidence is consistent with a phonological account in which nothing happens to the nasal in any of the circumstances in (9), without requiring an unquestioning acceptance that 'gestures' are phonological. The phonology may simply be indifferent to mechanical place assimilations of objects which are anyway without place specification.[20]

Furthermore, the orthodox elemental account of these assimilations only succeeds for examples (a) (b) and (d). [U] could be said to spread in the first two cases, and (d) retains the 'citation' (unspecified) shape. Example (e), by these lights, would require the spread of the normally inert [@] element, and if there is anything elemental going on in example (c) it is certainly not immediately obvious what it is.

The paradigm in (9) leads on to the question of how best to decompose dentals generally, and the English dental fricatives in particular. The minimal triplets 'fought/ thought /sought' (and the rather less felicitous 'Venn/ then/ zen') look, at first sight, as if they demand a straightforward melodic contrast involving different elements, or possibly an isomeric distinction. It might be logically possible to do this, but there is another solution which may be worth considering.

Some intra-coronal contrasts are more transparently available to melodic analysis than others. English [ʃ]/[s] involves the presence or absence (respectively) of element [I]. What to do with [s]/[θ] is not so immediately obvious. English coronal stops are generally alveolar, while Spanish (for instance) are generally dental.[21] For these languages, whatever melodic decomposition of the English alveolars or the Spanish dentals may be invoked, a phonological representation that gives proper deference to universals should take account of the equivalence of these (phonetically close) realisations within the contrastive inventory of each language. A coronal is a coronal: /t/ is /t/ in both languages.

---

[20] Cohn (1993) makes a comparable case for the non-phonological nature of English vowel nasalisation from physical evidence: she argues that the measurable 'gradient' rather than 'categorical' (slow rather than fast) realisation of this process reveals it to be 'phonetic'.

[21] The fact that [t] does contrast with [t̪] in Panjabi is typologically a relatively exotic characteristic, and one which, in the spirit of the present discussion, would need to be evaluated in the light of other facets of the language.

The use of a dental/alveolar contrast for fricatives is a distinction employed in both English and Spanish, whereas this distinction is relatively unusual in western European languages as a whole, being significantly absent from Portuguese, French, Italian and Greek, and also from German. Focussing on English, there are three other factors which show the dentals, rather than the alveolars, to be exotic. Speakers of dialects which utlise [θ/ð] acquire them in production notoriously late in life (maybe as late as six or more years of age). Before this time the adult forms may be realised as [f] or [t] (for the fortis) or [v] or [d] (for the lenis). Also, there is a parallel with the [f/v] realisation in other English dialects, where 'fought' and 'thought' are homophonic: [fɔːt]. Finally, the dentals are spectrally nearly- identical to the labiodentals. In acoustic analysis, it has been demonstrated that the principal physical correlate of the English lexical contrast is often durational, /f/v/ being uniformly longer than /θ/ð/.

The question is why should speakers bother to acquire such a fugitive distinction, and at a developmental stage which is far later than is normal for segmental contrasts? The answer could be that the pronunciation of English dentals is not part of language-dedicated phonology at all, but part of the 'human group mark system', whose existence is propounded in Kaye (1996). Other social primates possess such a system, so this idea broadly aligns us with our evolutionary cousins: where we differ, according to Kaye, is that ours is exclusively a vocal system. He goes on to advance the notion from ecological considerations that what atrophies at puberty in the human psyche is not the ability to acquire language, but the plasticity of this ability to 'conform'. If this is so, the distinctive vocalisations needed to qualify as a group member may be acquired by unconscious mimicry, and wouldn't need to be assigned to phonology at all. English dentals would then demand a different analysis to (say) Spanish dentals. They become a sociological variation on a labi(odent)al theme.

Of course English dentals do have a lexical function, and it could reasonably be proposed that we are overlooking this. What happens if we reject the 'sociological' analysis, and try to incorporate these objects into English phonology? The special factors affecting their acquisition and distribution must then be ignored, and what arguably remains true from the acoustic evidence of their primarily durational contrast with labiodentals is that they are still not properly termed coronal.

The crux of this argument is that we may need to take into account the possibility of two extra-phonological factors being responsible for some aspects of a given utterance. Blind mechanics may make things look more highly specified than they are, and we may also be making socially-inspired noises in parallel with linguistically generated ones. The first of these *caveats* is acknowledged to be better formulated than the second, but the extreme lateness of the acquisition of certain linguistic contrasts exemplified by English dental fricatives, if it were shown to be outside a phonological 'critical period', may ultimately prove to be good evidence of their separate function. Awareness of the cohabitation of mechanics and mimicry with phonology in utterances could make some of the murkier corners of melodic theory, and the theory of melodic acquisition, a little clearer.

6.5 The correlation of perception and production in acquisition

Is it possible to link the ideas presented here to infant production data? For instance, the ubiquity of CV utterances during the stage of pre-word productions is well attested (Jakobson 1962, Menn 1980, *et al.*), and it could be interesting to speculate, in terms of the present characterisation of the nature of CV, as to what the phonological content of these utterances is. It seems there are two logical options. Given that 'naked' prosody may be interpreted (§5.6, p.169), children may either be producing a nuclear domain empty of melodic content, or attaching maximally contrastive material from the melodic store to this syllabic template. The first option would seem to be a reasonable interpretation of a [mə] or [pə] utterance, and the second of a [ma] or [pa]. But maybe we need to temper our interpretations of pre-word utterances with considerations linked to the physiological immaturity of the producer.

In this context, a word is in order about the time-lag between perception and production. The perceptual experiments (both original and cited) reported in earlier chapters show that phonological acquisition well before the so-called 'babbling' stage has reached a level of development which includes a knowledge of far more than simply a canonical syllable, even if, as we have discussed, systems during year one of life are as yet unaligned. The simplest account of the time-lag is that non-linguistic knowledge of appropriate sensorimotor instructions must also be acquired early in life before successful articulations can be achieved. Phonetic evidence of the gradual acquisition of articulatory competence is presented by Bond *et al.*(1982). In this longitudinal study of one child, F1 and F2 measurements of articulations of known vowel targets are compared at different times within the age range 1:5:0 to 2:5:0. The results show that the areas of the vowel space used to express the different 'phones' become slowly more distinct, with gradually less overlap between the targets. Only at the end of the experimental period do such quintessentially contrastive vowels as [i] and [a] achieve discrete acoustic interpretations: the vowel targets are consistently moving apart, indicating that the child progressively becomes more capable of achieving precise aims.

Another factor which unequivocally does cause a time lag is the physiological maturation that must take place before the vocal tract can articulate any human speech. The larynx of a neonate is high in the throat, and the tract is wider in relation to its length than that of an adult: 'more similar to that of the adult chimpanzee than it is to that of the adult human' (Ingram 1989: 97). Only at six months old does the shape of the tract begin to conform at all to its eventual proportions, and Fletcher (1973) proposes that it may take as long as eight years after birth before physiological development allows fully co-ordinated articulatory patterns. Such lengthy processes of maturation are typical of general human development: as Gould pointed out (1980), the human neonate compared to any other mammal is an independently living fetus. It is born about nine months earlier than would be expected by analogy with other animals, both to allow its comparatively large cranium through the cervix, and also, arguably, to stimulate its neuropsychological ontogeny as soon as possible by exposing it to the environment.

The slow development of articulatory function, therefore, together with the need in the first months of life to establish neuromuscular control, must account for some of the time lag between perception and production. Including these factors in an acquisitional account means that reading phonological implications off pre-word production data has to be a less reliable endeavour than analysing perception tests. It appears that firm answers to questions about babbling (such as the one proposed in this section about the phonology of early CV) are for these reasons not available.

6.6 Chronology

A broad summation of the chronological stages of the acquisition of phonology via perception as proposed in this thesis is given in the following list. Ages are of course approximate. These stages are consistent with all the research reported in chapter 2, the original experimental data of chapter 4, and all theoretical considerations advanced here about the nature of the phonological faculty as a whole.

(i) ***Prosodic structure****: pre-natal - 1:0:0*

The mapping of psychoacoustic cues onto some prosodic structure can begin before birth, as this information is carried at low frequencies which are transmissible through intrauterine fluid. Normalisation is already utilised at around the time of birth, and some language-specific structure related to longer stretches of speech is being perceived at (at the latest) four days of age. Phonological word- and foot-structure begins to be perceived between six and nine months, and there is a general trend towards the perception of progressively less extensive domains.

(ii) ***Melodic primes****: birth - 1:0:0*

Once the organism is in the air, language-specific phonological acquisition of melodic primes can begin to take place selectively from the store of psychoacoustic factory-settings. Both categorical and prototypical strategies are utilised, and between 0:0:6 and 1:0:0 the discrimination of non-native categories/prototypes atrophies as perceptual space ossifies.

(iii) ***Integration:*** *1:0:0 and older*

The two systems align themselves, which constitutes a separate acquisitional task. Neither system need yet be fully mature. It is possible, for instance, that the child needs information from the integrated systems to parametrise the constituents of syllable structure.[22]

6.7 Summary

This chapter has sought to contextualise enqiries into early melodic acquisition which use the theoretical armoury wielded in the earlier chapters. A reduced elemental inventory has been set in a psychological frame (§6.2), and the potential for acquisitional variation in the assignation of this reduced set to psychoacoustic cues exemplified and discussed (§6.3). Section 6.4 sounded a brief caution against proposing phonological specifications and analyses where other systems may be 'interfering' in the speech signal, and in §6.5 reasons of physiological immaturity were cited for concentrating our attentions on infant perception, rather then production, in any investigation into early language acquisition. A chronological frame for phonological acquisition was also proposed (§6.6). It is hoped that further research in this area will prove these contextualisations to be both accurate and useful.

---

[22] This may become more obvious as our understanding of the 'reflections' of prosody and melody increases. Some recent work tends towards limiting prosodic structure universally to an ON domain comprising a single CV sequence (van der Hulst 1994, Lowenstamm 1996, Rennison 1998). If this idea is accepted, then the acquisition of what were branching structures becomes the acquisition of the knowledge of where inter-onset nuclei are licensed to be empty. In the more orthodox model, it is feasible that the perception of melodic material *in situ* prosodically is needed to set a parameter for a branching constituent as ON.

Chapter 7

Recapitulation and Answers

7.1 Recapitulation

The foregoing chapters have the stated aim of convincing the reader of the truth of the notion that phonological significance can be assigned to infants' perceptual abilities. Chapter 1 offered a definitional framework within which we could consider this matter by delimiting the nature of the phonological abstractions and perceptual strategies that seem to be relevant to this end. The independence of prosody and melody is a cornerstone of our understanding of phonology, and this was exploited in chapter 2 as a means of untangling some confusions that arise if infant discriminatory abilities are *not* considered to be phonological. In particular, the assertion was made that the development of prosody and melody and the alignment of these two systems must represent separate acquisitional tasks, and that this removes some contradictions inherent in the findings reported in the literature. In the terms of the thesis presented here, some of these findings betray the acquisition of a sensitivity to the phonation contrasts which map to elements [H] and [L].

Chapter 3 offers confirmation of the results reported by Kuhl, Bosch, and other researchers which have led to the characterisation of vowel perception as 'prototypical', and the proposal by Kuhl of the existence of NLMs. This chapter lays the groundwork for possible future testing by analysing the two front mid vowels of Catalan in terms of their melodic complexity, and reporting experimental results which argue that the prototypes for vowels of differing complexity are perceived by adult native speakers (and not by controls). Future tests on children of different ages may eventually illuminate this putative relationship between chronology and complexity, and so forge a link between early acquisition and phonological theory.

One of the prime motivations of the present work is to show that the importation of theory into perception testing opens up diverse opportunities. Chapter 4, then, turns to an investigation into the alignment of phonology and early perceptual abilities based on psychoacoustic diversity. The putative early acquisition of [H] and [L] is a relevant area of enquiry because of the phonetically and psychoacoustically disparate nature of the manifestations of these elements in onsets and nuclei.

Using empirical tests, a potential chronological parallel between the development of the phonative and tonal *alter ego*s of [H] and [L] is explored. Children have previously been shown to have language-specific sensitivity to phonation at six months (as demonstrated by research reviewed in chapter 2), and this sensitivity is exactly concordant with adult perceptions. The results of original experiments using Yorùbá adults and infants, and English controls, now show that there is a parallel perception of lexical tone by all members of the target Yorùbá groups, outside of a wider prosodic context. In contrast, the English adult results show non-linguistic perception of pitch, and English babies appear to be quite indifferent to lexical tonal cues.[1]

An anomalous inability on the part of the Yorùbá adults to perceive the presumed pitch cue for [L] was also present in the baby tests. Though there *may* be non-phonological explanations for this result on this particular occasion, there are enough typological asymmetries in the tonology of west African languages for it to be worth seeking a linguistic account. Chapter 5 proposed that no extant account of Yorùbá gives any explanation of these asymmetries, and the following solution was offered: the only nuclear element utilised by the language is [H], mapping to high tone; the perception of low pitch or a fall, hitherto said to cue 'low tone', maps to a prosodic juncture. Mid-tone is (as usual) non-existent from a phonological point of view. Nothing in either chapter 4 or chapter 5 is inconsistent with the notion that by six months of age a child has developed the tone perception relevant to its native language, just as it has for the perception of phonation. The tonology of Yorùbá can be seen to align with both adult and infant perception.

---

[1] Methodological *caveats* abound: see §4.4 and §4.5.

Attention was then focused on contextualising this enquiry into early acquisition: in chapter 6, psychological and chronological frames were made explicit. In addition, the notions of variability of elemental assignment, the non-specificity of melodic objects, and non-phonological speech sounds were discussed. These three factors may all be seen as conspiring to support reductionism in phonological theory.

7.2 Answers

In §1.2, (p.19), the following questions were framed (they are repeated here as (1) for convenience):

(1)

(a) Can the lights of phonology illuminate existing acquisitional findings, and do existing acquisitional findings integrate with phonological theory, especially with regard to the prosody/melody bifurcation?

(b) Is it possible to design infant acquisitional tests to answer properly formulated, specifically phonological, questions?

(c) Can we deliver an account of the acquisition of melodic primes?

(d) Do babies display evidence of language-specific phonological processing, such that:

  (i) parallel acquisition of items that are phonetically diverse, yet phonologically identical, may be demonstrated?

  (ii) there is any traceable connection between phonological complexity and acquisitional chronology?

Broadly, all these questions have been answered in the affirmative. As regards the first and second, the application of specifically phonological questions to child and infant perception testing can open up fresh lines of inquiry for developmental research, and once we start thinking this way, the formulation of questions and the design of relevant tests can be comparatively easy.

The work reported in chapters 4 and 5 concentrates on answering question (1(d(i))) (above) by example. The parallel acquisition of phonetically disparate, yet phonologically identical objects has been demonstrated in this instance. Question (1(d(ii))) remains open, but the (preliminary) tests reported in chapter 3 at least show one way in which it may one day be resolved. Because it is possible to answer the two questions in (1(d)), even by example, question (1(c)) can also be answered in the affirmative: we can deliver an account of the acquisition of melodic primes if we remain aware of the phonetic disguises in which they may show up, and the perceptual correlates that are relevant to them. Indeed, if we could not answer this question positively, that would be good reason to be wary about the theoretical base for melody that we have espoused.

7.3 Phonology and future perceptual research

It is fair to assess the value of the approach advanced in this thesis by seeing how readily it may be applied to research beyond the confines of the present enquiry. A program which is currently getting under way at the Department of Experimental Psychology, Oxford, is investigating, among other things, the precursors of children's acquisition of inflectional endings (Plunkett & Bryant 1998). They note that one-year-olds 'discriminate monosyllabic words that match in their vowels and initial consonants' but 'fail to distinguish monosyllabic words that rhyme ... from controls' and that this 'apparent attention to the initial segments of words, but not the final segments of words, poses a problem for the early recognition of inflectional categories' (op. cit.: 4). The 'words' they have in mind are of CVC construction ('fet, fem, kad, vik' etc.). Now English babies of this age will be acquiring English foot-structure, and in our terms, words of this type are left-headed trochees with a defective dependent syllable. It would be better to characterise their 'attention' as attention to prosodic form, rather than to 'segments'. Unless this phonological information is incorporated into the program, it cannot possibly occur to the researchers to investigate the relative perception of, say, [fet] and [ˈfetə], or [ˈpɔbə], [ˈpɔdə] and [ˈpɔgə], a putatively fruitful line of inquiry. The acquisition of (the phonological system for) English inflectional endings must postdate both the acquisition of the [ON] setting for the final empty-nucleus parameter and the loosening of constraints on items licensed in the rightmost onset (for instance, it is permitted not to harmonise dorsals).[2]

Phonologically informed perception testing represents one step forward in answering the need identified by another psychologist, Karmiloff-Smith, for future research to be focused explicitly on the notion 'that early domain-specific competences are directly linked to later capacities' (1995: 1308).

7.4 Conclusion

Nothing in any of the foregoing chapters is inconsistent with the maximally simple notion of phonological acquisition proposed at the start of this discussion. Characterising phonology as an abstract language-dedicated pattern-recognition system has allowed us to make a start on the investigation of an early developmental stage, which is to serve as an example for future work. The fact that it has been possible to link the primes and constituents of GP to acquisitional perception testing gives support to the theory as a whole. Phonological significance should be ascribed to infant perception.

This is one virtual bridge across the Euston Road (§1.1, p.15). It is only a small one, and doesn't take much of a payload yet, but it is hoped that more substantial structures will follow.

---

[2] See also the discussion of consonant harmony in §6.3.3, (pp.188-192).

Appendix A

An inventory of ten elements and their rough acoustic correlates.

The following is freely adapted and much abridged from Choo & Huckvale (1993).

Attached directly under root node:

[?]	'occluded', or 'edge'	(abrupt decrease in amplitude).
[h]	'noise'	(high frequency aperiodic energy).
[N]	'nasal'	(low frequency resonant peak).

Laryngeal (source) elements:

[H]	'high tone'	(in sonorant: rise in Fx) (in obstruent: high frequency aperiodic energy).
[L]	'low tone'	(in sonorant: fall in Fx) (in obstruent: continuous Fx).

Place (resonance) elements:

[A]	'mass'	(one energy peak in centre of frequency range).
[I]	'dip'	(low energy in centre of range; higher either side).
[U]	'rump'	(high energy at low frequency, falling off as frequency climbs).
[R]	'coronal'	(formant transitions associated with coronality).
[@]	'neutral'	(no acoustic property).

PAGE
MISSING
IN
ORIGINAL

Appendix B (i)

Edited 'Turner' log of Yorùbá infant perception test

```
Turner: Version 1.0
(c)1995 Mark Huckvale University College London
Date: Thu Jul 24 14:56:36 1997
Subject: YORUBA INFANT
Voter: lucy
Experimenter: phil
NumTrial: 10
StimCnt: 4
StimDur: 2.0s
VoteDur: 4.0s
ExpireDur: 8.0s
ReinforceDur: 5.0s
----------------------
15:38:49.4: Loaded items 1.01, 1.02 from xyotr.sfs
10 OR 12 TRAINING RUNS: APPEARED SUCCESSFUL

15:52:50.4: Loaded items 1.01, 1.02 from x1820.sfs
15:53:16.6: Start run of 10 trials using x1820.sfs stimuli
15:55:17.1: Trials=6 Hits=0/3 FalseAlarms=0/3
180/200 NO

15:56:58.7: Start run of 10 trials using x1719.sfs stimuli
15:57:04.4: Start trial 1 = Control stimulus
15:57:23.7: Start trial 2 = Control stimulus
15:57:50.6: Start trial 3 = Test stimulus
15:57:51.6: Play Stimulus A
15:57:52.2: Experimenter signal
15:57:53.5: Voter signal
15:57:53.5: Double vote
15:57:53.5: Reinforcer on
15:57:53.7: Play Stimulus B
15:57:54.0: Voter signal
15:57:54.0: Double vote
15:57:54.0: Reinforcer on
15:57:55.7: Play Stimulus B
15:57:57.7: Play Stimulus B
15:57:59.0: Reinforcer off
15:58:10.5: Trials=3 Hits=1/1 FalseAlarms=0/2
PAUSE
15:58:13.0: Loaded items 1.01, 1.02 from x1719.sfs
15:58:36.4: Start run of 10 trials using x1719.sfs stimuli
15:58:39.7: Start trial 1 = Control stimulus
15:58:53.9: Start trial 2 = Test stimulus
15:59:07.9: Start trial 3 = Control stimulus
15:59:19.7: Start trial 4 = Test stimulus
15:59:38.8: Trials=4 Hits=0/2 FalseAlarms=0/2
(= 7 1/3, 0/4, i.e. NOT SIGNIFICANT)
170 /190 NO

16:00:18.6: Loaded items 1.01, 1.02 from xyotr.sfs
RETRAINED: APPEARED SUCCESSFUL

16:01:22.9: Loaded items 1.01, 1.02 from x1921.sfs
16:01:26.5: Start run of 10 trials using x1921.sfs stimuli
16:01:33.4: Start trial 1 = Control stimulus
16:01:56.3: Start trial 2 = Control stimulus
16:02:12.3: Start trial 3 = Control stimulus
16:02:27.4: Start trial 4 = Test stimulus
16:02:31.5: Experimenter signal
16:02:33.0: Voter signal
16:02:33.0: Double vote
16:02:33.0: Reinforcer on
16:02:43.3: Start trial 5 = Test stimulus
16:02:51.4: Experimenter signal
16:02:52.0: Play Stimulus B
16:02:52.0: Voter signal
16:02:52.0: Double vote
16:02:52.0: Reinforcer on
16:03:00.6: Start trial 6 = Control stimulus
16:03:23.2: Start trial 7 = Test stimulus
16:03:24.5: Play Stimulus A
16:03:26.5: Play Stimulus B
16:03:28.5: Voter signal
16:03:29.6: Experimenter signal
16:03:29.6: Double vote
16:03:29.6 Reinforcer on
16:03:41.3: Start trial 8 = Test stimulus
```

```
16:03:42.8: Play Stimulus A
16:03:44.8: Play Stimulus B
16:03:46.8: Play Stimulus B
16:03:48.9: Play Stimulus B
16:03:50.9: Play Stimulus B
16:03:52.9: Play Stimulus A
16:03:55.0: Play Stimulus A
16:03:57.0: Play Stimulus A
16:03:59.0: Play Stimulus A
16:04:04.6: Start trial 9 = Control stimulus
16:04:17.9: Start trial 10 = Test stimulus
16:04:22.7: Voter signal
16:04:23.2: Experimenter signal
16:04:23.2: Double vote
16:04:23.2: Reinforcer on
16:04:53.5: Pause on
16:05:02.0: Pause off
16:05:06.9: Trials=10 Hits=4/5 FalseAlarms=0/5
190 /210 SIGNIFICANT

16:05:25.5: Loaded items 1.01, 1.02 from x1416.sfs
16:05:28.3: Start run of 10 trials using x1416.sfs stimuli
16:05:31.1: Start trial 1 = Test stimulus
16:05:45.7: Start trial 2 = Control stimulus
16:05:59.9: Start trial 3 = Control stimulus
16:06:15.6: Start trial 4 = Test stimulus
16:06:17.1: Play Stimulus A
16:06:19.0: Experimenter signal
16:06:19.1: Voter signal
16:06:19.1: Double vote
16:06:19.1: Reinforcer on
16:06:24.8: Reinforcer off
16:06:31.9: Start trial 5 = Control stimulus
16:06:35.0: Voter signal
16:06:35.2: Experimenter signal
16:06:35.2: Double vote
16:06:35.2: Reinforcer on
16:06:40.8: Reinforcer off
16:06:54.2: Start trial 6 = Control stimulus
16:06:58.5: Experimenter signal
16:07:00.2: Voter signal
16:07:00.2: Double vote
16:07:00.2: Reinforcer on
16:07:29.7: Start trial 7 = Control stimulus
16:07:58.2: Start trial 8 = Test stimulus
16:08:25.7: Start trial 9 = Control stimulus
16:08:53.3: Start trial 10 = Test stimulus
16:08:55.3: Voter signal
16:08:56.3: Experimenter signal
16:08:56.3: Double vote
16:08:56.3: Reinforcer on
16:09:10.6: Trials=10 Hits=2/5 FalseAlarms=2/5
140/160 A HINT OF SOMETHING BUT NOT SIGNIFICANT

16:09:28.0: Loaded items 1.01, 1.02 from x2123.sfs
16:09:34.8: Start run of 10 trials using x2123.sfs stimuli
16:09:39.5: Start trial 1 = Test stimulus
16:09:52.2: Start trial 2 = Control stimulus
16:10:08.3: Start trial 3 = Control stimulus
16:10:20.6: Start trial 4 = Test stimulus
16:10:39.3: Start trial 5 = Control stimulus
16:10:54.6: Trials=5 Hits=0/2 FalseAlarms=0/3
210 /230 NO (?)

PAUSE

16:15:18.9: Loaded items 1.01, 1.02 from x1920.sfs
16:15:22.7: Start run of 10 trials using x1920.sfs stimuli
16:15:28.3: Start trial 1 = Control stimulus
16:15:35.5: Experimenter signal
16:15:42.8: Start trial 2 = Control stimulus
16:15:54.1: Start trial 3 = Control stimulus
16:15:58.3: Trials=3 Hits=0/0 FalseAlarms=0/3
190/200 NO:10Hz OR BORED?

16:16:21.3: Start run of 10 trials using x2223.sfs stimuli
16:16:23.9: Start trial 1 = Control stimulus
16:16:48.8: Start trial 2 = Test stimulus
16:16:50.3: Experimenter signal
16:16:55.4: Start trial 3 = Control stimulus
16:17:08.6: Start trial 4 = Test stimulus
16:17:17.1: Start trial 5 = Control stimulus
16:17:27.8: Start trial 6 = Control stimulus
16:17:39.3: Trials=6 Hits=0/2 FalseAlarms=0/4
220/230 NO. HIGH IN Fx RANGE SHOULD PREJUDICE

16:18:33.8: Reinforcer on
16:18:38.8: Reinforcer off
16:18:42.5: Reinforcer on
16:18:47.5: Reinforcer off

16:27:18.7: Done
```

Appendix B (ii)

Edited 'Turner' log of Catalan adult perception test

The first two letters of a stimulus file refer to the prototype to which they relate: EI = [e], and EA = [ɛ]. The two numbers refer to the positons of the pair of stimuli on a particular vector, 0 being the prototype itself, and 1,2,3 and 4 being progressively further out by 30 mels at a time. File names use the same coding, with the addition of X to identify a test file, and one or two letters after the prototype ID to indicate which vector they are sited on. Thus file XEANW23 has a pair of stimuli relating to the [ɛ] prototype taken from its 'northwest' vector respectively 60 and 90 mels distant from the centre.

```
Turner: Version 1.0
(c)1995 Mark Huckvale University College London
Date: Thu May  2 16:14:23 1996
Subject: CATALAN ADULT
Voter: eva
Experimenter: phil
NumTrial: 10
STIM: CAT: E:A34 A23 A01 A03 A12 A02 A13 I34 I02 I24 I23 I01 I14 I13
StimCnt: 4
StimDur: 2.0s
VoteDur: 4.0s
ExpireDur: 8.0s
ReinforceDur: 5.0s
----------------------
16:15:49.8: Loaded items 1.01, 1.02 from xeae04.sfs
16:17:32.3: Start run of 10 trials using xeae04.sfs stimuli
16:18:37.8: Trials=10 Hits=5/5 FalseAlarms=0/5
EA04 PERCEIVED CORRECTLY (AS TEST)

16:18:55.4: Loaded items 1.01, 1.02 from xeae34.sfs
16:19:03.8: Loaded items 1.01, 1.02 from xeae34.sfs
16:19:09.3: Start run of 10 trials using xeae34.sfs stimuli
16:20:44.6: Trials=10 Hits=0/5 FalseAlarms=0/5
EA34 0%

16:21:08.4: Loaded items 1.01, 1.02 from xeae24.sfs
16:21:12.2: Start run of 10 trials using xeae24.sfs stimuli
16:24:16.3: Trials=10 Hits=5/5 FalseAlarms=0/5
EA24 100% CORRECT

16:24:45.4: Loaded items 1.01, 1.02 from xeae01.sfs
16:24:51.0: Start run of 10 trials using xeae01.sfs stimuli
16:26:20.2: Trials=10 Hits=0/5 FalseAlarms=0/5
EA01 0%

16:26:37.1: Loaded items 1.01, 1.02 from xeae13.sfs
16:30:10.5: Trials=10 Hits=5/5 FalseAlarms=0/5
EA13 100% CORRECT

16:32:28.1: Loaded items 1.01, 1.02 from xeae23.sfs
16:32:30.4: Start run of 10 trials using xeae23.sfs stimuli
16:34:41.0: Trials=10 Hits=5/5 FalseAlarms=0/5
EA23 100% CORRECT

16:35:17.4: Loaded items 1.01, 1.02 from xeae02.sfs
16:35:23.1: Start run of 10 trials using xeae02.sfs stimuli
16:37:41.0: Trials=10 Hits=5/5 FalseAlarms=0/5
EA02 100% CORRECT

16:38:04.7: Loaded items 1.01, 1.02 from xeae12.sfs
16:38:07.1: Start run of 10 trials using xeae12.sfs stimuli
16:40:26.7: Trials=10 Hits=5/5 FalseAlarms=0/5
EA12 100% CORRECT

16:40:50.0: Loaded items 1.01, 1.02 from xeae01.sfs
16:40:54.2: Start run of 10 trials using xeae01.sfs stimuli
16:43:26.4: Trials=10 Hits=5/5 FalseAlarms=0/5
EA01 100% CORRECT

16:45:29.6: Loaded items 1.01, 1.02 from xeae34.sfs
16:45:32.2: Start run of 10 trials using xeae34.sfs stimuli
16:47:46.9: Trials=10 Hits=4/5 FalseAlarms=0/5
EA34 SIGNIFICANT

16:56:42.4: Loaded items 1.01, 1.02 from xeie34.sfs
16:56:51.5: Loaded items 1.01, 1.02 from xeie34.sfs
16:56:54.6: Start run of 10 trials using xeie34.sfs stimuli
16:59:18.7: Trials=10 Hits=4/5 FalseAlarms=0/5
EI34 SIGNIFICANT

17:00:00.7: Loaded items 1.01, 1.02 from xeie12.sfs
17:00:04.5: Start run of 10 trials using xeie12.sfs stimuli
17:02:29.6: Trials=10 Hits=0/5 FalseAlarms=0/5
EI12 0%

17:02:58.1: Loaded items 1.01, 1.02 from xeie01.sfs
17:03:04.0: Start run of 10 trials using xeie01.sfs stimuli
17:05:10.5: Trials=10 Hits=0/5 FalseAlarms=0/5
EI01 0%

17:05:38.7: Loaded items 1.01, 1.02 from xeie14.sfs
17:05:41.8: Start run of 10 trials using xeie14.sfs stimuli
17:07:52.4: Trials=10 Hits=5/5 FalseAlarms=0/5
EI14 PERCEIVED CORRECTLY (AS TEST)

17:08:15.6: Loaded items 1.01, 1.02 from xeie24.sfs
17:08:19.0: Start run of 10 trials using xeie24.sfs stimuli
17:10:41.0: Trials=10 Hits=5/5 FalseAlarms=0/5
EI24 100% CORRECT

17:10:50.4: Loaded items 1.01, 1.02 from xeie02.sfs
17:10:55.6: Start run of 10 trials using xeie02.sfs stimuli
17:13:06.1: Trials=10 Hits=4/5 FalseAlarms=0/5
EI02 SIGNIFICANT

17:13:23.1: Loaded items 1.01, 1.02 from xeie23.sfs
17:13:26.5: Start run of 10 trials using xeie23.sfs stimuli
17:15:40.6: Trials=10 Hits=4/5 FalseAlarms=1/5
EI23 SIGNIFICANT

17:16:08.9: Loaded items 1.01, 1.02 from xeie01.sfs
17:16:16.6: Start run of 10 trials using xeie01.sfs stimuli
17:18:37.3: Trials=10 Hits=3/5 FalseAlarms=1/5
EI01 NOT SIGNIFICANT

17:18:52.1: Loaded items 1.01, 1.02 from xeie34.sfs
```

```
17:18:55.7: Start run of 10 trials using xeie34.sfs stimuli
17:21:12.4: Trials=10 Hits=4/5 FalseAlarms=0/5
EI34 SIGNIFICANT

17:22:43.5: Done
```

Appendix C

Responses logged during infant perception testing for pairs of Yorùbá stimuli differentiated by pitch only

(i) Yorùbá infants

	Stimuli (Hz)	*Total*	*Hits*	*False Alarms*
Child 1	180/200	10	0/5	0/5
	190/210	10	3/5	0/5
	220/240	7	0/4	0/3
Child 2	140/160	7	1/4	2/3
	170/190	7	0/3	1/4
	190/210	10	3/5	0/5
	210/230	4	0/1	1/3
	220/240	10	3/5	1/5
Child 3	160/180	6	0/3	0/3
	190/210	10	5/5	0/5
	210/230	7	0/4	0/3
Child 4	170/190	8	0/4	0/4
	190/210	10	4/5	0/5
	210/230	9	0/4	1/5
Child 5	140/160	10	2/5	2/5
	170/190	7	1/3	0/4
	180/200	6	0/3	0/3
	190/210	10	4/5	0/5
	210/230	5	0/2	0/3
Child 6	140/160	10	2/5	1/5
	180/200	6	0/3	0/3
	190/210	7	0/3	1/4
	210/230	7	1/4	2/3

(ii) English infants

Child 7	160/180	4	0/2	0/2
	170/190	10	0/5	1/5
	190/210	10	1/5	0/5
Child 8	190/210	9	2/5	0/4

Child 9	180/200	8	1/5	0/3
	190/210	7	0/4	0/3
	210/230	4	2/3	0/1
Child 10	190/210	5	0/2	0/3
Child 11	170/190	2	0/2	0/0
	190/210	10	0/5	0/5
	210/230	10	1/5	1/5
Child 12	140/160	10	2/5	1/5
	180/200	7	0/3	0/4
	190/210	8	1/4	0/4
	210/230	10	2/5	0/5

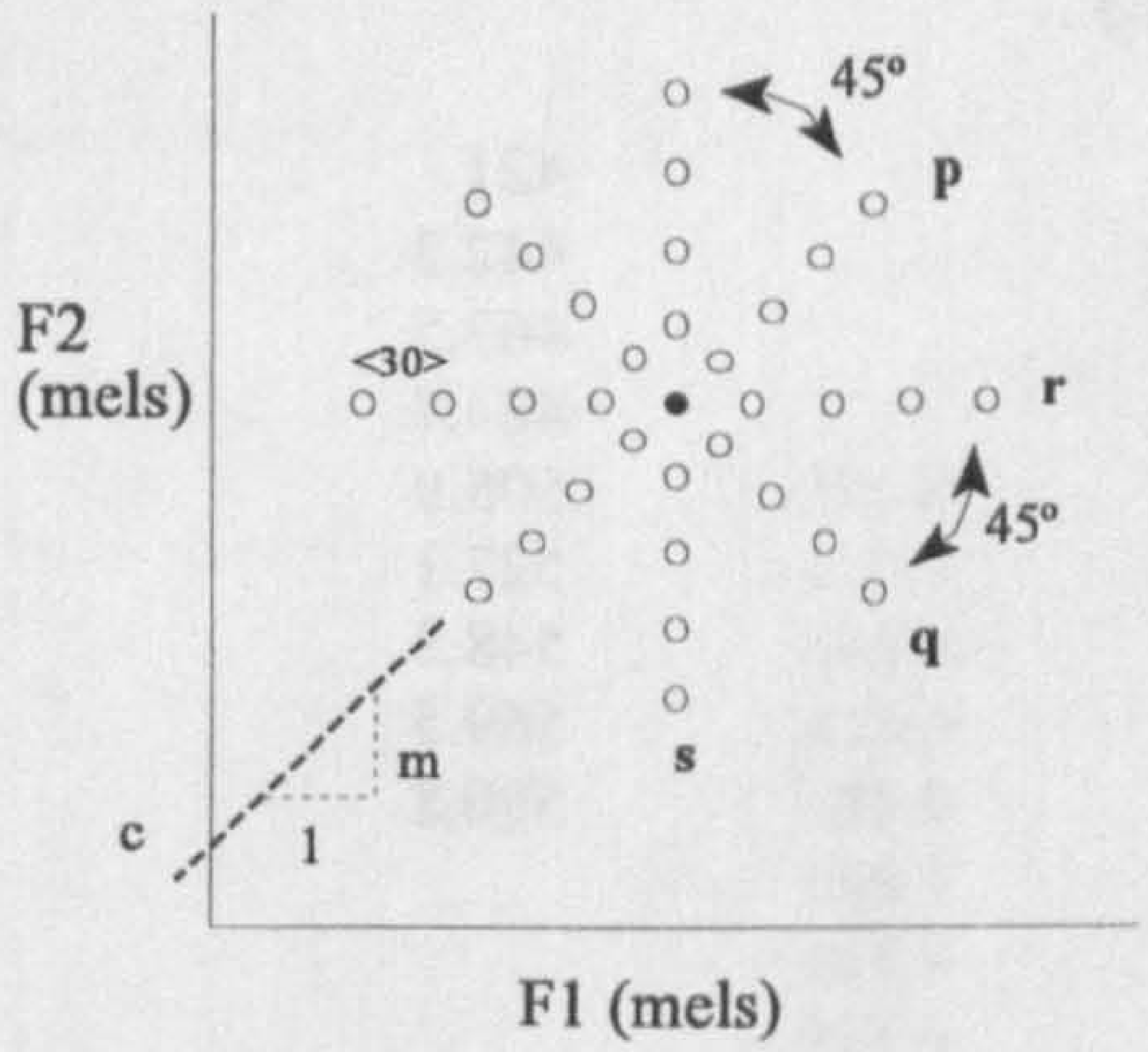
Appendix D

F2 vs F1 in Hz and mels for Catalan front mid-vowels

Given two central points for (F1 ; F2) of (420 ; 2580) (for [e]) and (570 ; 2500) (for [ɛ]) in Hz, use equation (A) below to convert these to mels:

(A) $$\text{mel} = \frac{1000}{\ln(2)} \times \ln\left(1 + \frac{\text{Hz}}{1000}\right)$$

		Hz:	mels:
[e]	F1	420	505.9
	F2	2580	1840
[ɛ]	F1	570	650.8
	F2	2500	1807.4



F2 vs F1 plotted at 30 mels equidistant intervals around a central point (505.9 ; 1840) or (650.8 ; 1807.4)

*Equation for vector **p**:*

m = slope
c = intercept on F2 axis
F2 = m.F1 + c

$$m = \frac{\Delta F2}{\Delta F1} = \frac{30 \sin 45°}{30 \cos 45°} = \frac{30 / \text{sqrt } 2}{30 / \text{sqrt } 2} = 1$$

Using central point to find c:

for [e]: $1840 = 1 \times 505.9 + c$

$c = 1334.1$

$\therefore F2 = F1 + 1334.1$

for /ɛ/: 1807.4 = 1 x 650.88 + c

c = 1156.6

∴ F2 = F1 + 1156.6

*Equation for vector **q**:*

for line at - 45°, m = - 1

Using central point to find c:

for /e/: 1840 = -1 x 505.9 + c

c = 2345.9

∴ F2 = - F1 + 2345.9

for /ɛ/: 1807.4 = -1 x 650.88 + c

c = 2458.2

∴ F2 = - F1 + 2458.2

For vectors ***p*** and ***q***, values of F1 are calculated by adding or subtracting 30/sqrt(2) from the central point. The values of F2 are calculated from F1 using the equations delivered above.

For /e/:

F1 values for vectors ***p*** and ***q*** .

$F1_j = 505.9 + j.\frac{30}{\text{sqrt2}}$

421
442.3
463.5
484.7
505.9
527.1
548.3
569.5
590.8

F2 values for vector ***p***

$F2_j = F1_j + 1334.1$

1755.1
1776.4
1797.6
1818.8
1840
1861.2
1882.4
1903.6
1924.9

F2 values for vector ***q***, by F2 = F1 + 2345.9, will render the same values as F2 for vector ***p***, but in inverse order.

For /ɛ/:

F1 values for vectors ***p*** and ***q*** .

$F1_j = 650.8 + j.\frac{30}{\text{sqrt2}}$

565.9
587.2
608.4

629.6
650.8
672
693.2
714.4
735.7

F2 values for vector $\boldsymbol{p}$

$F2_j = F1_j + 1156.6$

1722.5
1743.8
1765
1786.2
1807.4
1828.6
1849.8
1871
1892.3

F2 values for vector $\boldsymbol{q}$, by $F2 = F1 + 2458.2$, will render the same values as F2 for vector $\boldsymbol{p}$, but in inverse order.

For vector $\boldsymbol{r}$,

For [e]:

F2 = 1840, and

$F1_j = 505.9 + 30j$

385.9
415.9
445.9
475.9
505.9
535.9
565.9
595.9
625.9

For [ɛ]:

F2 = 1807.4, and

$F1_j = 650.8 + 30j$

530.8
560.8
590.8
620.8
650.8
680.8
710.8
740.8
770.8

For vector $\boldsymbol{s}$,

For [e]:

F1 = 505.9 and

$F2_j = 1840 + 30j$

1720
1750
1780
1810
1840
1870
1900
1930
1960

For [ɛ]:

F1 = 650.8 and

$F2_j = 1807.4 + 30j$

1687.4
1717.4
1747.4
1777.4
1807.4
1837.4
1867.4
1897.4
1927.4

The values of F1 and F2 for vectors ***p q r s*** make up the plot required in mels around each prototype.

To convert these plots to units of Hz, rearrange equation (A) to deliver equation (B):

(B) $$\text{Hz} = 1000\text{x}\ [e\ (\frac{\ln(2).\text{mel}}{1000})- 1]$$

Using equation (B) for [e]:

F2 for vector ***r*** = 2580.1

F1 for vector ***s*** = 420

F1 vectors ***p*** and ***q***	F2 vector ***p***	F2 vector ***q***	F1 vector ***r***	F2 vector ***s***
338.9	2375.6	2797	306.7	2294.4
358.7	2425.6	2741.6	334.1	2363.6
378.8	2476.4	2686.9	362.2	2434,3
399.3	2527.8	2633.1	390.8	2506.4
420	2580.1	2580.1	420	2580.1
441	2633.1	2527.8	449.8	2655.3
462.4	2686.9	2476.4	480.3	2732.1
484	2741.6	2425.6	511.4	2810.6

506	2797	2375.6	543.2	2890.6

Using equation (B) for [ε]:

F2 for vector ***r*** = 2500.1

F1 for vector ***s*** = 570

F1 vectors ***p*** and ***q***	F2 vector ***p***	F2 vector ***q***	F1 vector ***r***	F2 vector ***s***
480.4	2300.2	2712.1	444.7	2220.8
502.3	2349.1	2658	475.1	2288.4
524.5	2398.7	2604.6	506.1	2357.5
547.1	2449	2552	537.7	2428.1
570	2500.1	2500.1	570	2500.1
593.3	2552	2449	603	2573.7
616.9	2604.6	2398.7	636.7	2648.7
640.8	2658	2349.1	671.1	2725.4
665.2	2712.1	2300.2	706.2	2803.7

PAGE
MISSING
IN
ORIGINAL

References

Akinlabi A. (1993). Underspecification and the Phonology of Yoruba /r/. *Linguistic Inquiry* 24, 139-160

Anderson J.M. & Jones C. (1974). Three theses concerning phonological representations. *Journal of Linguistics* 10, 1-26

Anderson J.M. & Ewen C.J. (1987). *Principles of Dependency Phonology* Cambridge: CUP

Anderson S.R. (1978). Tone features. In *Tone: a Linguistic Anthology* (V.Fromkin, editor) New York: Academic Press

Anderson S.R. (1981). Why phonology isn't 'natural'. *Linguistic Inquiry* 12, 493-539

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

Aslin R.N. & Pisoni D.B. (1980b). Some developmental processes in speech perception. In *Child Phonology Vol. 2. Perception* (G. Yeni-Komshian, J.Kavanagh, & C.Ferguson, editors) New York: Academic Press

Backley P. & Takahashi T. (1998). Element activation. In *Structure and Interpretation: Studies in Phonology* (E. Cyran, editor) Lublin: Folium

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

Baker M. (1988). *Incorporation* Chicago: University of Chicago Press

Bamgbose A. (1966). *A Grammar of Yoruba* Cambridge: CUP

Barrett S.E. (1997). *Prototypes in speech perception* PhD dissertation. University of Cambridge

Barrett S.E. (1998). The speech prototype and its role in infant speech perception development. Paper presented at *UCL Linguistics Seminar*, UCL, London, March 1998.

Bertoncini J., Bijeljac-Babic R., Jusczyk P.W., Kennedy L.J., & Mehler J. (1988). An investigation of young infants' perceptual representations of spech sounds. *Journal of Experimental Psychology: General* 117, 21-33

Beckman M.E. & Pierrehumbert J. (1986). Intonational structure in English and Japanese. *Phonology Yearbook* 3, 255-310

Best C.T. (1991). Phonetic influences on the perception of non-native speech contrasts by 6-8 and 10-12 month olds. *Biennial meeting of the Society for Research in Child Development*: Seattle WA.

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

Bond Z.S., Petrosino L. & Dean C.R. (1982). The emergence of vowels: 17 to 26 months. *Journal of Phonetics* 10, 417-422

Bosch L., Costa S., & Sebastian N. (1994). La estructura interna de las categorías fonéticas: Percepción de vocales e identificación de prototipos en Español y Catalán. In *Nuevos Horizontes de la Lingüística Aplicada* (J. Llisterri & D. Poch, editors). ms., Universitat Autònoma de Barcelona

Bright W. (1957) *The Karok Language* Los Angeles: University of California Press

Brockhaus W., Ingleby M. & Chalfont C. (1996). *Acoustic signatures of phonological primes* ms., University of Manchester

Brody M. (1997). *Mirror Theory* ms., UCL London

Browman C.P. & Goldstein L. (1989) Articulatory gestures as phonological units. *Phonology* 6, 201-251

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

Burton-Roberts N.& Carr P.(1996). On Speech and Natural Language. *Newcastle & Durham Working Papers in Linguistics* 4, 13-40

Cabrera-Abreu M. (1996). *Intonation without [L]* PhD dissertation. UCL, London

Carr P. (1991). Lexical properties of post-lexical rules: Postlexical derived environment and the Elsewhere Condition. *Lingua* 85, 41-54

Charette M. & Göksel A. (1998). Licensing constraints and vowel harmony in Turkic languages. In *Structure and Interpretation: Studies in Phonology* (E. Cyran, editor) Lublin: Folium

Cho Y-m.Y. (1988). Korean Assimilation. *Proceedings of the Seventh West Coast Conference on Formal Linguistics* 41-52

Chomsky N. (1986). *Knowledge of Language* New York: Praeger

Chomsky N. (1988). *Language and Problems of Knowledge* Cambridge, MA.: MIT Press

Choo W. & Huckvale M. (1993). Perceptual confusions as evidence for non-linear phonology. *UCL Working Papers in Linguistics* 5, 325-354

Chumbow B.S. (1982). Contraction and Tone Polarization in Ogori. *Journal of West African Languages* 12, 89-103

Clements G.N. (1983). The hierarchical representation of tone features. In *Current Approaches to African Linguistics Vol 1* (I.R. Dihoff, editor) Dordrecht: Foris

Clements G.N. & Keyser S.J. (1983). *CV Phonology: a Generative Theory of the Syllable* Cambridge, MA: MIT

Cobb M. (1995). Vowel harmony in Zulu and Basque: The interaction of licensing constraints, H-licensing, and constituent structure. *SOAS Working Papers in Linguistics* 5, 23-40

Cohn A.C. (1993). Nasalisation in English: phonology or phonetics? *Phonology* 10, 43-81

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

Corina A. & Sandler W. (1993). On the nature of phonological structure in sign language. *Phonology* 10, 165-208

Connell B. & Ladd D.R. (1990). Aspects of pitch realisation in Yoruba. *Phonology* 7, 1-29

Coulter G.R. & Anderson S.R. (1993). Introduction. In *Phonetics and Phonology Vol.3 Current Issues in ASL* (G.R. Coulter, editor) London: Academic Press

Crystal D. (1991). *A Dictionary of Linguistics and Phonetics* (3rd. edition). Oxford: Blackwell

Cutler A. & Mehler J. (1993). The periodicity bias. *Journal of Phonetics* 21, 103-108

Dehaene-Lambertz G. & Dehaene S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature* 370, 292-295

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

Demuth K. (1993). Issues in the acquisition of the Sesotho tonal system. *Journal of Child Language* 20, 275-301

Dennett D.C. (1995). *Darwin's Dangerous Idea* London: Penguin

Disner S.F. (1986). On describing vowel quality. In *Experimental Phonology* (J.J. Ohala & J.J. Jaeger, editors) New York: Academic Press

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

Fant G. (1973). *Speech Sounds and Features* Cambridge, MA: MIT

Fifer W.P. & Moon C. (1989). Psychobiology of newborn auditory preferences. *Seminars in Perinatology* 13, 430-433

Fletcher S.G. (1973). Maturation of the speech mechanism. *Folia Phoniatrica* 25, 161-172

Foltin R. (1996). *The right edge of German verbs: phonology and automatic speech recognition* ms., University of Vienna

Goldsmith J.A. (1990). *Autosegmental and Metrical Phonology* Oxford: Blackwell

Gould S.J. (1980). *Ever Since Darwin* London: Penguin

Grieser D. & Kuhl P.K. (1989). Categorisation of speech by infants: support for speech-sound prototypes. *Developmental Psychology* 25, 577-588

Halle M. & Clements G.N. (1983). *Problem Book in Phonology* Cambridge, MA: MIT

Halle M. & Stevens K.N. (1971). A note on laryngeal features. *Quarterly Progress report of the Research Laboratory of Electronics (MIT)* 101, 198-213

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

Harris J. (1996). Phonological output is redundancy-free and fully interpretable. *UCL Working Papers in Linguistics* 8, 551-574

Harris J. (1997). Licensing Inheritance: an integrated theory of neutralisation. *Phonology* 14, 315-370

Harris J. & Lindsey G. (1993). There is no level of systematic phonetic representation. *UCL Working Papers in Linguistics* 5, 355-373

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. (1994). *The alignment of morphological constituents with those of syntax and phonology.* ms., UCL, London

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

Hayward K., Watkins J., & Oyètádé A. (1998). *The Phonetic Interpretation of Register: Evidence from Yorùbá.* ms., SOAS, London

Hollenbach B. (1984). *The phonology and morphology of tone and laryngeals in Copala Trique* PhD dissertation. University of Arizona

House D. (1990). *Tonal Perception in Speech* Lund, Sweden: LUP

Hulst H.G. van der (1989). Atoms of segmental structure: components, gestures and dependency. *Phonology* 6, 253-284

Hulst H.G. van der (1993). Units in the analysis of signs. *Phonology* 10, 209-241

Hulst H.G. van der (1994). Radical CV Phonology: the locational gesture. *UCL Working Papers in Linguistics* 6, 439-477

Hulst H.G. van der (1996). Symmetry in Phonology. Paper presented at the *Fourth Phonology Meeting*, University of Manchester, 1996.

Hyman L.M. (1975). *Phonology; Theory and Analysis* New York: Holt, Rinehart & Winston

Hyman L.M.(1986). The representation of multiple tone heights. In *The Phonological Representation of Suprasegmentals* (K. Bogers, H. van der Hulst & M. Mous, editors) Dordrecht: Foris

Ingram D. (1989). *First Language Acquisition* Cambridge: CUP

Itô J. & Mester A. *Word binarity and weak layering* ms., University of California, Santa Cruz, CA

Jackendoff R. (1987). *Consciousness and the Computational Mind* Cambridge, MA: MIT

Jakobson R. (1962). Why Mama and Papa? In *Selected Writings* The Hague: Mouton

Jusczyk P.W. (1985). The high amplitude sucking procedure as a methodological tool in speech perception research. In *Measurement of audition and vision in the first year of postnatal life: a methodological overview* (G, Gottlieb & N.A.Krasnegor, editors) Norwood, NJ: Ablex

Jusczyk P.W. (1993). From general to language-specific capacities: the WRAPSA model of how speech perception develops. *Journal of Phonetics* 21, 3-28

Jusczyk P.W., Cutler A., & Redanz N.J. (1993). Infants' preferences for the predominant stress patterns of English words. *Child Development* 64, 673-687

Jusczyk P.W. & Derrah C. (1987). Representations of speech sounds by young infants. *Developmental Psychology* 23, 648-654

Kaisse E. M. & S. Hargus (1993). Introduction. In *Phonetics and Phonology vol. 4* (S. Hargus & E. Kaisse, editors) San Diego: Academic Press

Karmiloff-Smith A. (1995). Annotation: the extraordinary cognitive journey from fetus through infancy. *Journal of Child Psychology and Psychiatry* 36, 1293-1313

Katamba F. (1998). Cophonologies: The Case of Luganda. Paper presented at the *Sixth Phonology Meeting*, University of Manchester, May 1998.

Kaye J. (1990). 'Coda' licensing. *Phonology* 7, 301-330

Kaye J. (1996). Sex, lies, primates and Trevor McDonald. Paper presented at the *SOAS Linguistics Departmental Seminar*, London, March 1996.

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

Kent R.D. & Read C. (1992). *The Acoustic Analysis of Speech* San Diego, CA: Whurr

Kimura D. (1961). Cerebral dominance in the perception of verbal stimuli. *Canadian Journal of Psychology* 15, 166-171

Kimura D. (1964). Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology* 16, 355-358

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. (1980). Perceptual constancy for speech-sound categories in early infancy. In *Child Phonology Vol. 2. Perception* (G. Yeni-Komshian, J.Kavanagh, & C.Ferguson, editors) New York: Academic Press

Kuhl P.K. (1986). Theoretical contributions of tests on animals to the special-mechanisms debate in speech. *Experimental Biology* 45, 233-265

Kuhl P.K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories: monkeys do not. *Perception and Psychophysics* 50, 93-107

Kuhl P.K. (1992). Infants' perception and representation of speech: development of a new theory. *ICSLP Proceedings*: Banff, Canada

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. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology* 4, 812-822

Kuhl P.K. & Meltzoff A.N. (1982). The bimodal perception of speech in infancy. *Science* 218, 1138-1141

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

Kuhl P.K. & Miller J.D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception and Psychophysics* 3, 279-292

Kuhl P.K., Williams K.A., Lacerda F., Stevens, K.N., & Lindblom B. (1992). Linguistic experience alters phonetic perception in infants by six months of age. *Science* 255, 606-608

Ladd D.R. (1996). *Intonational Phonology* Cambridge: CUP

Ladefoged P. & Maddieson I. (1996). *The Sounds of the World's Languages* Oxford: Blackwell

Lass R. (1984). *Phonology: an introduction to basic concepts* Cambridge: CUP

Lecanuet J. (1993). Fetal responsiveness to voice and speech. In *Developmental Neurocognition: speech and face processing in the first year of life* (B. de Boysson-Bardies, P. Jusczyk, P. MacNeilage, J. Morton & S. de Schonen, editors) Dordrecht: Kluwer Academic

Levelt W.J.M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition* 42, 1-22

Levin J. (1985). *A Metrical Theory of Syllabicity* PhD dissertation. Cambridge, MA: MIT

Lieber R. (1992). *Deconstructing Morphology* Chicago: University of Chicago Press

Lindblom B. (1986). Phonetic universals in vowel systems. In *Experimental Phonology* (J.J. Ohala & J.J. Jaeger, editors) New York: Academic Press

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

Lowenstamm J. (1996). CV as the only syllable type. In *Current Trends in Phonology: Models and Methods vol. 2* (J. Durand and B. Laks, editors) Salford: European Studies Research Institute

McCarthy J.J. & Prince A. (1986). *Prosodic Morphology* ms., University of Massachusetts

McNeill D. (1970). *The Acquisition of Language* New York: Harper & Row

Maddieson I. (1984). *Patterns of Sounds* Cambridge: CUP

Marr D. (1982). *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information* San Francisco: W.H.Freeman

Matisoff J.A. (1973). Tonogenesis in southeast Asia. In *Consonant Types and Tone* (L Hyman, editor) Los Angeles: University of Southern California Linguistics Program.

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

Menn L. (1980). Phonological theory and child phonology. In *Child Phonology Vol. 1. Production* (G. Yeni-Komshian, J.Kavanagh, & C.Ferguson, editors) New York: Academic Press

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

Nasukawa K. (1998). An integrated approach to nasality and voicing. In *Structure and Interpretation: Studies in Phonology* (E. Cyran, editor) Lublin: Folium

Neeleman A. & Weerman F. (1997). L1 and L2 Word Order Acquisition. *Language Acquisition,* 6.2, 125-170

O'Connor J.D. & Arnold G.F. (1973). *The Intonation of Colloquial English* (2nd. edition) London: Longman

Ohala J.J. (1997). The relation between phonetics and phonology. In *The Handbook of Phonetic Sciences* (Hardcastle & Laver, editors) Oxford: Blackwell

Odden D. (1995). Tone: African languages. In *The Handbook of Phonological Theory* (J. Goldsmith, editor) Oxford: Blackwell

Ola O. (1993a). *Benue-Congo Prosodic Phonology and Morphology in Optimality Theory* PhD dissertation. University of Tulane

Ola O. (1993b). Non-exhaustive syllabification in Yoruba. Paper presented at the *24th Annual Conference in African Linguistics*, Ohio State University, Columbus, OH

Ola O. (1996). *Properheadedness and Binarity: Prosodic Words in Yoruba* ms., University of British Columbia

Oyètádé B.A. (1988). *Issues in Yoruba tone* PhD dissertation. SOAS, London.

Palmada F.B. (1991). *La fonologia del català i els pricipis actius* PhD dissertation. Univertitat Autònoma de Barcelona

Paradis C. & Prunet J-F. (editors) (1991). *Phonetics and Phonology 2: The special status of coronals: internal and external evidence* , San Diego,CA: Academic Press

Peters A. (1973). A new formalisation of downdrift. *Studies in African Linguistics* 4, 139-154

Piaget J. & Inhelder B. (1969). *The Psychology of the Child* London: Routledge & Kegan Paul

Pierrehumbert J. (1980). *The Phonology and Phonetics of English Intonation* PhD dissertation. MIT

Pierrehumbert J. & Beckman M.E. (1988). *Japanese tone structure* Cambridge, MA: MIT press

Plunkett K. & Bryant P. (1998). *Phonological Processes in Early Language Acquisition* Research proposal, Department of Experimental Psychology, Oxford.

Poser W. (1982). Phonological representations and Action-at-a-distance. In *The Structure of Phonological Representations* (H. van der Hulst & N. Smith, editors) Dordrecht: Foris

Price G. (1991). *An Introduction to French Pronunciation* Oxford: Blackwell

Pulleyblank D. (1986). *Tone in Lexical Phonology* Dordrecht: Reidel

Pulleyblank D. (1988). Vocalic Underspecification in Yoruba. *Linguistic Inquiry* 19, 233-270

Rennison J. (1990). On the elements of phonological representations: the evidence from vowel systems and vowel processes. *Folia Linguistica* 24, 175-244

Rennison J. (1998). Contour segments without subsegmental structures. In *Structure and Interpretation: Studies in Phonology and Syntax* (E. Cyran, editor) Lublin: Folium

Rice K. (1996). Default variability: the coronal-velar relationship. *Natural Language and Linguistic Theory* 14, 493-543

Ritter N. (1996). The role of asymmetrical headedness in the reduction theory of elements. Paper presented at the *19th GLOW colloquium,* University of Athens, April 1996

Ritter N. (1997). Headedness as a means of encoding phonological stricture. In *Phonology in Progress - Progress in Phonology* (G. Booij & J. van der Weijer, editors) The Hague: Holland Academic Graphics

Rivera-Gaxiola M. (1998). *Electrophysical correlates of cross-linguistic speech perception in adult native English speakers* PhD dissertation. UCL, London

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

Rowlands E.C. (1969). *Yoruba* New York: Hodder & Stoughton

Schane S. (1984). The fundamentals of particle phonology. *Phonology Yearbook* 1, 129-155

Selkirk E.O. (1986). On Derived Domains in Sentence Phonology. *Phonology Yearbook* 3, 271-405

Siertsema B. (1959). Problems of Phonetic Interpretation II: Long Vowels in a Tone Language. *Lingua* 8, 42-64

Slater A. (1992). The visual constancies in early infancy. *The Irish Journal of Psychology* 13, 411-424

Smith N.V. (1973). *The Acquisition of Phonology* London: CUP

Spencer A. (1991). *Morphological Theory* Oxford: Blackwell

Stahlke H. (1974). The development of the three-way tonal contrast in Yoruba. In *Proceedings of the Third Annual Conference on African Linguistics* (E. Voeltz, editor) Bloomington: Indiana University Press

Stemberger J.P. (1983). *Speech errors and theoretical phonology: a review* Bloomington: Indiana Linguistics Club

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

Stuart-Smith J. (1996). The representation of a tonal contrast in British Panjabi. Paper presented at the *Fourth Phonology Meeting*, University of Manchester, May 1996.

Summerfield Q. (1983). Audio-visual speech perception, lipreading and artificial stimulation. In *Hearing Science and Hearing Disorders* (Haggard & Lutman, editors) London: Academic Press

Takahashi T. (1993). A farewell to constituency. *UCL Working Papers in Linguistics* 5, 375-410

Takahashi T, (1994). Constraint interaction in Aranda stress. *UCL Working Papers in Linguistics* 6, 479-507

Tsushima T., Takizawa O., Sasaki M., Shiraki S., Nishi K., Konho M., Menyuk P. & Best C. (1994) Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: language-specific developmental changes in speech perception abilities. *ICSLP 94 Proceedings* Yokohama

Uyechi L. (1996). *The Geometry of Visual Phonology* CSLI: Stanford, CA.

Walker C. (1995). ATR harmony as h-licensing. *SOAS Working Papers in Linguistics and Phonetics* 5, 108-120

Werker J. & Lalonde C.E. (1988). Cross-language speech perception: initial capabilities and developmental change. *Developmental Psychology* 24, 672-683

Werker J. & Polka L. (1993). Developmental changes in speech perception: new challenges and new directions. *Journal of Phonetics* 21, 83-102

Woods D.L. & Elmasian R. (1986). The habituation of event-related potentials to speech sounds and tones. *Electroencephalography and Clinical Neurophysiology* 65, 447-459

Yip M.J.W. (1980). *The Tonal Phonology of Chinese* PhD dissertation. MIT.

Yoshida Y. (1995). *On pitch accent phenomena in Standard Japanese* PhD dissertation. SOAS, London

Zsiga E.C. (1992). A mismatch between morphological and prosodic domains in evidence for two Igbo rules. *Phonology* 9, 101-135