Phonology and infants' perceptual abilities: asking the right questions^{*}

PHIL HARRISON

1 Introduction

A good deal of painstaking empirical research is detailed in the literature on language acquisition in the first year of life. The majority of it, for obvious reasons, focuses 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 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 a direction for future 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 has done previously (Menn 1980), that Roman Jacobson, 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.

We begin with a brief outline of the phonological theory we are adopting. Then follows some exemplification of the problems arising from the lack of integration of these phonological notions into acquisitional research, with a couple of theoretical proposals which could potentially clarify things, and finally we discuss the broad shape of a research program which seems to emerge spontaneously from these considerations, and propose the first step in that program.

2 Theoretical background

2.0 Following a train of argumentation traceable through McCarthy (1979), Clements & Keyser (1983), Harris (1983), Clements & Goldsmith (1984) and Goldsmith

^{*}I am very grateful to John Harris for his helpful comments on an earlier draft of this paper. All remaining defects are entirely my fault.

(1990), we accept the independence and interdependence of melodic and prosodic representations: this idea is very briefly fleshed out in the following subsections.

2.1 A melodic theory

The model adopted here is that developed by Kaye, Lowenstamm and others (Kaye, Lowenstamm & Vergnaud 1985, Schane 1984 *inter alia*), and specifically as expounded by Harris (Harris 1994, Harris & Lindsey 1993). Melodic material is analyzable into a small set of 'elements', or primes, 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 have been used to account for some distinctions previously regarded as 'phonemic'. The linking of elements to prosodic structure has been said to underlie processes such as spreading and assimilation, and their delinking to account for lenition trajectories and diphthongisation.



Figure 1. Elements and infrasegmental geometry.

Elements are affected as classes by phonological processes, and infrasegmental geometry is therefore as robust a tenet in element theory as it is in a feature-theoretic approach (such as that employed in Clements 1985). The nine elements which are the maximum at present generally considered necessary to generate melodic expressions

are set out in figure 1, under the node to which they associate. (For the acoustic correlates of these elements, see Choo & Huckvale 1993 p.341.)

[R], associated with coronality, is not included here for reasons which Backley has previously set out in detail (Backley 1993). A further element often referred to in the literature, [ATR], is also missing from this inventory. This element has been thought necessary to account for the contrast between 'tense and 'lax' vowels, but it is putatively possible to achieve this by positing complexes headed by [@]. This element 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. So we can argue that in English /u/ = [U] and /u/ = [U@]. The existence of [ATR] as an element is also undermined by its distinctly un-elemental property of being phonetically unrealisable¹.

Segments conventionally regarded as falling on both sides of the consonant/vowel bifurcation share resonance elements. We cannot, therefore, on purely melodic grounds, make any principled distinction between them. It may be pretheoretically argued that the elements which correspond with opposite ends of the sonority hierarchy at least offer some kind of polarity; so that stops, which contain [?], are 'quintessentially' consonantal, and that [A] is the 'voweliest' vowel, these two epitomising the CV contrast. But even if it is true that [?] and [A] are the least likely elements to combine, all we can formally say is that they are unlikely to surface in the same prosodic position as each other, for reasons as yet unspecified. It is uncontroversial that [I] and [U] may surface (as semivowels) in 'consonantal'² positions, and Broadbent (Broadbent 1991) has argued that [A] does so too, its phonetic realisation being /r/-liaison. It may seem at first more questionable that [?] is realisable in a 'vowel' position, but the uncontentious proposal that nasals and/or laterals may be 'syllabic' would require this to be so.

By contrast with the elusiveness of the CV contrast here, there is a more tenable dichotomy between sonorants and obstruents, solely on the grounds of the autosegmental melodic theory discussed here. When they are part of an obstruent, the elements under the laryngeal node, [H] and [L], signal phonation contrasts, but in

¹Kaye *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' (p.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.

 $^{^2 {\}rm I}$ am being deliberately vague about 'syllabic' positions here, as I don't wish to presuppose any particular prosodic theory.

sonorants they signal tonal contrasts. The use of any two unary elements delivers a potential four-way paradigm, so empirical support for this notion is derived from the neat account it gives of a language such as Gujarati, where a four way 'voicing' contrast exists.

However, it does mean we have to accept that in non-tonal languages with no fullyvoiced 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; in time to come, such a proposal may overcome this particular objection.

One vital contribution of this melodic theory for the purposes of the present discussion is that phonological processes are accounted for without recourse to any formal definition of 'phoneme': as an informal sign for a concatenation of melodic components, it retains a useful notational purpose, but that is all.

2.2 Prosodic constituents and licensing

The first caveat when trying to derive a sensible prosodic theory is to avoid alphabetic chauvinism. The knowledge that when Japanese native speakers hear the line of haiku <shinshin-to> there are clearly five, and not three, 'syllables', and that when Amharic is written in Ge'ez script the sequence [pit] is 'syllabified' as [pi#t] ought to prevent us from allowing the syllable the totemic standing which it is often accorded.

It is empirically true that the configuration conventionally written as 'CV' occupies a uniquely unmarked status in natural languages. All of the world's investigated languages allow CV sequences, and some, such as Bantu languages, allow no extension of the headed nuclear domain other than to append a single initial consonant. The data available from studies of pre-word stage children shows that the majority of proto-word utterances are CV, further supporting this notion. At a later stage, it is well attested that children will reduce more complex adult forms to CV sequences. And although the intuitions of native speakers of such languages as Japanese, referred to in the preceding paragraph, have supported the proposal of subsyllabic morae to account for the quantitative structure of such a language, over 60% of all morae turn out to be CV (Otake *et al.* 1993).

However, in section 2.1 above we argued that the consonant/vowel distinction was not formally supportable as a definition of melodic content. Rather we argued that the sonorant/obstruent bifurcation may be linguistically justified, aligning with the fact that the elements [H] and [L] have different acoustic realisations when part of an obstruent than when they are part of a sonorant.

So we need to align the empirical ubiquity of CV with our inability to characterise CV melodically, leaving us only the option of defining it prosodically. But some of the theoretical proposals of the past decade or so have allowed us to do without a 'CV' tier altogether and argued instead for a purely skeletal level, which is independent of melodic content. An early nail was driven into the coffin of the CV tier by James Harris (Harris 1983) by demonstrating that the processes which assign stress target rhymes: the advent of the Onset/Rhyme level makes the CV tier redundant and requires the proposal of a pure timing level, either skeletal or moraic. Levin (1985) showed that during the process of compensatory lengthening a VC sequence may become VV, and if we persist with a CV tier, vocalic material will become attached to a C slot. Kaye & Lowenstamm (1985) support this argument using data from Tiberian Hebrew. This language has geminate consonants, but in a context where gemination is expected, /?/ fails to do so. Instead the preceding vowel spreads to the target timing slot. The skeleton allows this process without ambiguity.

Figure 2. Formal constituents within the nuclear domain.

The 'syllable peak' is therefore now formally characterised as the supraskeletal nucleus, and the nucleus is the head of the phonological domain. All other segmental positions must be licensed, either directly, or via an intermediate position, by the head. Licensing relationships are local and directional. The projections of heads at higher prosodic levels are local for processes such as long-distance harmony. Within prosodic structure the only constituents which we have been able to empirically support are onset, rhyme, and nucleus.

Each of these constituents, within the theory under discussion, exhibits maximally binary branching. (For details and empirical support of these notions see Kaye Lowenstamm & Vergnaud 1990, and Harris 1994.)

The vitality of 'licensing' in this theory is underlined by the fact that the existence of even the three remaining constituents is under fire; Takahashi (1993) argues that licensing relations between heads, and projections of heads, will account for all prosodic structure.

Essential to the present discussion is that nowhere in the prosodic hierarchy is there either a 'syllable' or any level that could be termed 'CV'. These terms are thus formally

defunct from both a melodic and a prosodic point of view. Continued use of them in phonological discussions, unless they be defined in some new and particular way, will lead to confusion.

A system such as this which uses concepts like 'head-licensing' and 'projection' cries out for syntactic parallels, and these have been variously drawn. Levin (1985) likens nuclear projection to X-bar schema and Harris 1994 extends the principle that structure is preserved throughout derivation to both phonology and syntax, undermining any phonological accounts which appeal to 'resyllabification' and deriving a further parallel between syntactic empty categories and phonological empty positions. The existence of both domain-internal and domain-final empty nuclei, putatively occupied by a phonetically unrealised [@], allows the present prosodic theory to account for instances of elision and epenthesis in a number of contexts. Encrevé (1988) proposes the existence of empty onsets to deliver a theoretic account of the dichotomy between 'h-muet' and 'h-aspiré' in French; in this instance the unrealised element appears to be [?]. We simply note, for the moment, that the existence of empty positions is well-attested within this prosodic theory.

3 Relating early acquisitional strategies to phonology

3.0 In this section we discuss the sometimes apparently contradictory results achieved by researchers into phonetic acquisition, and propose that the integration of these results with our theoretical arguments may help unravel the contradictions.

3.1 Native-Language-Magnet theory

Many of the influential experiments in early acquisition (Eimas *et al.* 1971, Eilers, Gavin & Wilson 1979) concentrate on the perception of stops, with particular emphasis on the categorical perception of VOT³. In some more recent endeavours, attention has instead been focused on the acquisition of vocalic contrasts, but the findings argue in exactly the same direction as the more venerable VOT findings: there is an innate available repertoire of perceptual contrasts, which will be shaped by native language input.

Further, the *given* repertoire need not be language-dedicated, but arise as a result of general perception strategies also present in animals other than humans.

The results of experiments by Kuhl, (cited in Kuhl 1979, 1985, 1992, 1993, and Kuhl *et al.*, 1992) argue that infants of only six months of age perceive the contrasts

³We discuss the results of these experiments in section 4.

of their native language. These experiments show infants performing well above chance on several phonetic categorization tasks. A major focus of Kuhl's research has been the infant's perception of the vowel prototypes present in the language which he or she is acquiring. These prototypes can be quantified acoustically within a natural psychophysical area bounded by the first and second formants. Such results have led to the development by Kuhl of Native-Language-Magnet (henceforth NLM) theory.

Categorical perception of discrete areas within the psychoacoustic domain which can be defined by F1 and F2 is given at birth, and moreover is given to both humans and (at least) chinchillas (Kuhl 1986). We would expect from our common ancestry that some of our perceptions will be shared by other creatures, especially those close to us on the evolutionary scale, and that other abilities will be unique to, and universal within, our species. So we share with other mammals a genetic inheritance which impels perception of the map in figure 3 below. Though this map directly plots two psychoacoustic dimensions, it is entirely conceptual and could be drawn in any way that conveys the idea that at birth humans can discriminate all phonetic contrasts (Kuhl, personal correspondence). Even restricting ourselves to only these two dimensions among the several possible ones, the plot would presumably be just as valid with F1 on the ordinate and F2 minus F1 on the abscissa, so we should be careful not to posit an unnecessarily simplistic relationship between these representations and the parametrisation of the primes which can be active in structural relationships.

The fact that both babies and chinchillas categorically perceive this psychoacoustic space in an identical fashion does not diminish the unique nature of human ontogeny in its *reshaping* of this map in early life. NLM theory proposes that humans, and only humans, are able to develop a finite number of perceptual vowel targets, each of which will become located within one of the subdivisions of the above diagram. 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.



Figure 3. Innate psychophysical space (Kuhl 1992).

What of the responses to vowel distinctions which are genetically possible developments but are irrelevant for the infant's native language? NLM theory directly predicts that without reinforcement, these will atrophy: it is biologically commonplace now 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 (Best 1991, Werker and Lalonde 1988). So by the age of six months infants acquiring three different languages will have become responsive to the three contrastive vowel prototype inventories in figure 4. These diagrams are once again intended as conceptually informative; positions are only approximate.



Figure 4. Effects of linguistic environment at 6 months old (Kuhl 1992).



Figure 5. NLMs cause certain boundaries to disappear (Kuhl 1992).

Note that with the development of targets at particular values, some of the boundaries that divide the phonetic space become irrelevant to perception, and that the ability to perceive them at all will disappear: we can represent this state of affairs for our three hypothetical six-month-old language acquirers as in figure 5.

The presence of perceptual magnets for phonetic categories also actively interferes 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 proximity of the Japanese native /r/ to both English /r/ and /l/ explains why this should be a classically 'difficult' foreign distinction to acquire. Within the

phonological model of adult language we are adopting, the correlate of this acoustic distinction is the characterisation of /r/ and /l/ as differing by only one melodic element: the 'stop' element [?] is componential in /l/ but not in /r/.

This correlation raises the whole question of the mapping of these phonetic categories onto autonomous phonological entities. 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 cooperation with Meltzoff (Kuhl & Meltzoff 1979) 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 perception generally, I suggest that they may take place independently of the specific computations which relate NLMs to the primes of phonological representations. Just how this latter relation is accomplished is, as we have said, not at all clear, but my contention is that it is a valid question to ask, and that it can be asked validly only by being lucid about the entities we are trying to relate to each other.

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

Bearing in mind the perceptual finesse evinced at six months of age by Kuhl's work and proposed by her theory, we now question how this may tie in with the *size* of the segments perceived. There is a body of research which points to suprasegmental chunks as the units of early perception: this is exemplified in Jusczyk & Derrah (1987) and Mehler *et al.* (1988), among others. This idea potentially runs counter to Kuhl's theory: the two apparently divergent views are presented in two articles in Volume 21 of the *Journal of Phonetics* (Jusczyk 1993 and Kuhl 1993). Both are based firmly on empirical findings, and these empirical findings are at first sight contradictory. In the subsequent discussion we attempt to align the two sets of findings using a single theoretical assumption.

There is a good deal of research devoted to establishing the size of the first perceivable chunks of the acoustic signal. 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 is well supported.

Mehler *et al.* (1988) cite experiments which demonstrate that by the age of four months children can distinguish their own language from prosodic clues alone.

Removing everything above 400 Hz. from the speech signal by the use of a low-pass filter effectively masks all but the intonational and timing information, but when presented with such input, English infants can still tell English from French and Russian. To see if 'syllable-timing' versus 'stress-timing' could be significant, the experiment was repeated using pairs of syllable-timed languages (French and Italian), and stress-timed languages (English and Russian). This was found not to affect the infant's discriminatory ability.

Further 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 to *both* /bu/ and /du/. The same result has been shown in Bertoncini *et al.* (1988) for vowels: habituation to /bi, si, li, mi/ is followed by dishabituation to both /di/ and /da/. These findings have led to the conclusion 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. 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. Why should this be?

Kuhl (1993) argues that the part of this 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): using this technique, an animated toy animal in a plexiglass box requires a 45-degree head-turn by the infant to be visible, and is presented to the infant as reward for a 'correct' response. With this method, subjects can be reinforced for detecting similarities, and 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 or anyone else that I have read 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 prepared to do, and wants to retain it as' *a* unit of analysis'. 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 *et al.* (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 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 (1993) that though the subsequent development of such a contrastive repertoire may cause NLMs to 'repel' each other and further increase the psychoacoustic distance between them, the existence of NLMs at six months old is a result of 'assimilation' only. This is assimilation in the Piagetian definitional sense of 'filtering or modification of the input'.

3.3 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 *et al.* on older infants: if suprasegmental categories are only developed after six months, and 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' can only be achieved when perception has 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 me to be an obligation on the part of any theory of development to include some account of a continuum between these two, or at least an interface where one is predictably transduced into the other.

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. What we might suggest is that NLMs are evidence of the development of raw melodic contrasts, unattached to the prosodic hierarchy, which itself is developing independently from such perceptual clues as Jusczyk *et al.* have discerned.

If the independent and parallel development of prosody and melody were demonstrated, we would have the answer to the apparent anomalies between Kuhl and Jusczyk's conclusions. A clue to another 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. 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 NLMs 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.

A further prediction made by the proposal that NLMs are directly related to later representations springs to mind, and may bear empirical testing: some magnets should be stronger than others. If there is a direct relationship between phonological primes and NLMs, we may be able to demonstrate that those NLMs which relate to cross-linguistically less marked, more robust, targets develop earlier or more easily. In addition, it may be that 'uncoloured', or inactive, elements (for vowels, realisations of [@]), don't correlate with an NLM but are simply expressions of an otherwise unspecified constituent in a particular prosodic relationship. In the following section, I will suggest a way to start investigating such speculations.

In the meantime it is worth underlining 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. Since the 'phoneme' remains only as a shorthand term for a melodic construct, we cannot look to it as a *bona-fide* theoretical entity. Its definition, though, appears to be one basis for researchers to refuse to assign any phonological relevance to infant perception.

To help characterise an autonomous phonological level without phonemes, in the terms here intended, 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 have any idea where he is, even if the environment is a very familiar one⁴. 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.

It seems that a major barrier to the concept of an autonomous phonology as here intended is set up by the term 'phonemic perception', which ought not to be used as a synonym for 'phonological perception'. As we have previously stated, in phonological theory the phoneme has fallen apart into a set of non-linear primal abstracts which can be transparently accommodated at the level of representation which we are attempting to identify here. However, 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 last point may seem to some theoreticians an obvious one, but even the latest 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

⁴I have had enough conversations with friends about this state of consciousness to be convinced that it is not a personal pathology.

speech processing is a rapidly paced phenomenon which requires the listener to impose a phonemic identity on incoming sounds...' (Kent & Read 1992).

4 The direction of future research

How can we begin to harness these speculations into a sensible research program? If we can design perceptual tests that are able to illuminate patterns of mentation in infancy which foreshadow the active processes in adult language, or which are directly predictable from general theory, then we may establish some clues to the developmental process. In this section, we choose one controversial theoretical proposal, and outline an experiment which may shed some light upon it.

The traditional fortis/lenis distinction between English stop consonants is characterised in the terms developed in Government phonology and related theories (see, for instance, Kaye, Lowenstamm & Vergnaud 1990) 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 French, Greek and Spanish 'voiceless' stops. In these languages, the phonological distinction is rather between the neutral series and the 'fully voiced' set, which contains the element [L], whose presence is betrayed by the assimilation of voicing to adjacent positions, a phenomenon which never occurs from English lenis stops (see Harris 1994 pp.133-137). Now as we have discussed previously, the laryngeal elements [L] and [H] are proposed to be componential both in obstruents, where they give rise to phonation contrasts, and in nuclear positions where they are the active tonal elements in lexical tone languages. There are diachronic arguments for this identification: Hyman (1975) reports that what was once a fully-voiced obstruent in some southeast Asian languages has become at a later stage realised as a low tone on the following vowel.

We now need to recapitulate the results of two much-discussed experiments in infant categorical perception for stop consonants, those of Eimas *et al.* (1971) and those of Eilers, Gavin & Wilson (1979). The first of these investigations reports a universal VOT discriminatory boundary at around +25ms for bilabial stops in infants as young as one month old. As we know, however, input can alter such perception in the space of a few months, and the second experiment, a cross-linguistic study of 6 - 8 month old infants acquiring Spanish and English, which are languages that place categorical boundaries at different values of the VOT spectrum, reports that while the English children were more sensitive to the English system (where the boundary coincides with that of the (1971) experiment), the Spanish children were statistically

more able to perceive *both* this boundary and the native-language one (at around 0ms). This finding suggests that there is a species-universal tendency to discriminate categorically using a value of approximately +25ms. lag from release as the categorical boundary, but that given enough reinforcement from native-language input, perception of this distinction may be lost in favour of one elsewhere in the VOT continuum.

It should be noticed that the VOT value of the 'species-universal' boundary coincides with the most highly-coloured acoustic region in the continuum. A look at spectrographic analyses of these type of stimuli reveals that the 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 F1 transition is being progressively attenuated and the excitation of the upper formants is also altering. In this respect, then, the acquisition of phonology is 'unnatural' in the sense that Anderson (1981) uses the term with reference to the fully developed system and to diachronic change. Mechanical (auditory or articulatory) conditions do not determine a phonological system. Though 'naturalness' may certainly be one constraint on the development of such a system, it is in any case arguably extralinguistic. Given that we regard language as an autonomous cognitive faculty, we would not hypothesise that direct analogues would be found during the acquisition process any more than Anderson expects to find them in adult language. On the empirical side, even though we may trace certain phonetic 'jumping-off points' for some developmental processes, the language specific acquisition of stop contrasts by the Spanish and English infants that we have been considering ought to be evidence enough that this relationship is only one factor in the matter. We should also note, in this context, that the lack of acoustic salience of the lead (Spanish) contrast does not mark it out as in any way marginal: this contrast is widely attested as phonologically significant in a diverse set of languages.

We wish, then, to test the hypothesis that VOT perception at 6-8 months of age is already being transduced into phonological information having systemic import. If this is true, then the Spanish/English results indicate that [H] discrimination is present regardless of target language, and that [L] discrimination only arises if reinforced.

This asymmetry is certainly echoed in a typological skew found in adult languages. [H] is more generally found as an active tonal component of a nuclear position than is [L]. In Bantu languages, it is the only underlying tone, and its robustness in these languages is attested to by the process of the 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.

It would seem sensible, therefore, to extend these perceptual experiments to infants whose target languages are exemplars of both [H] and [H]/[L] systems to discover

ultimately if the same pattern is present in the acquisition of tonal contrast as has been demonstrated in the acquisition of VOT contrast. Such a test would have four logical potential outcomes:

- Statistical significance may be obtained for the perception of both contrasts. An initial result which suggested equal discrimination of both [H] and [L] type contrasts will not necessarily suggest anything theoretically significant about lexical tone discrimination: the perception of the pitch contour of the utterance, independent of any linguistic relevance, would itself account for such a result. But this result would suggest that we retest this discriminatory ability in older infants to try to establish when and how the percept of pitch becomes mapped onto lexical items and/or acquires intonational significance as distinct from its purely psychoacoustic interpretation.
- 2) No statistical significance may be demonstrated for the perception of either contrast. If at six months of age lexical tone is not discriminated at all, but VOT is discriminated (as has long been established), then it may be possible to propose that humans at six months are using the phonetic cues in phonation contrasts to map on to phonological primes but are not yet using acoustic pitch changes for this purpose. Even though perception of intonation patterns may be demonstrated in infants of this age, there may be some physical parameter of pitch (duration and/or frequency related) which renders it unavailable for the perception of segmental information until general strategies are further developed.⁵ A further possible interpretation of this result is that Kuhl was right to suggest that assimilation is the sole influence on perceptual development at this age. But it is not tenable that individual phonological systems magically spring, fully-armed, into being, and the onus would remain to account for the development of a pattern-recognition system from some phonetic/systemic interface.
- 3) A positive correlation may emerge between the pattern in the acquisition of phonation contrasts and that in the acquisition of tone. This would obviously be the most rewarding result, as it parallels both the VOT results and the patterns extant in the world's languages, and would certainly merit following up with tests using other tone languages.

 $^{^{5}}$ A speculative parallel with the late arrival of contrasts between non-strident fricatives in production may be drawn.

4) An inverse relation between the acquisition of phonation and tone could be demonstrated. This would run directly counter to the theoretical predictions: we'd be pretty sure we were on the wrong track.

5 A note on pitch and lexical tone

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

This observation raises a question for perceptual testing in general. If we can thus identify a norm for pitch perception independent of any structural context, and since we know that innately given abilities are adapted by the mind-brain when they are integrated to higher structural levels, can we ascertain whether any gain or loss in finesse of perception comes of the mapping of pitch stimuli onto higher levels? Such a gain or loss ought to be possible to trace in the perception of both speech and music. As far as infant development is concerned, we have here to account for mapping from an acoustic dynamic to a unary prime, but this by no means marks out the acquisition of tone as special, as an identical mapping takes place in the perception of stops. Lexical tone acquisition, though, may still conceivably turn out to be a separate issue developmentally from the acquisition of systemic contrasts from other segmental information. Clumeck's work has previously raised this possibility, mainly from production studies: he suggests (Clumeck 1980) that the acquisition of lexical tone acquisition of other segmental material but lags behind the acquisition of utterance sized intonational contours.

6 Summary

Any account of early acquisition should be locatable in the framework of a general phonological theory, and any general theory should be able to include an account of early acquisition. I suggest that the relationship of the two fields of research has largely been ignored and that unnecessary presumptions have been made about the significance (or lack of significance) of early phonetic perception to the acquisition of phonology: it should not be simply taken for granted that a particular phonetic focus has a phonological representational correlate, but we can derive inspiration from extant findings on infant perception to try to illuminate the developing interface between the two systems in early life. Importing theoretical notions, such as the independence of melody and prosody, into acquisitional research may help to clarify the processes of acquisition themselves.

References

Anderson S.R. 1981 Why phonology isn't "natural". Linguistic Inquiry 12, 493-539

- Backley P. 1993 Coronal: the undesirable element. In UCL Working Papers in Linguistics 5 London: University College
- 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
- 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.
- Broadbent J. 1991 Linking and intrusive /r/ in English. In UCL Working Papers in Linguistics 3 London: University College
- Choo W., & Huckvale M. 1993 Perceptual confusions as evidence for non-linear phonology. In UCL Working Papers in Linguistics 5 London: University College
- Clements G.N. 1985 The geometry of phonological features. Phonology Yearbook 2, 223-250
- Clements G.N. & Keyser S.J. 1983 CV Phonology: a Generative Theory of the Syllable Cambridge, MA: MIT
- Clements G.N. & Goldsmith J.A. 1984 Autosegmental Studies in Bantu Tone New York: Foris
- Clumeck H. 1980 The acquisition of tone. In *Child Phonology Vol. 1. Production* (G. Yeni-Komshian, J.Kavanagh, & C.Ferguson, editors) New York: Academic Press
- Demuth K. 1993 Issues in the acquisition of the Sesotho tonal system. *Journal of Child Language* **20**, 275-301
- 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
- Eilers R.E., Gavin W., & Wilson W.R., 1979 Linguistic experience and phonemic perception in infancy: a cross-linguistic study. *Child Development* **50**, 14-18

- Eimas P., Siqueland E., Jusczyk P.W. & Vigorito J. 1971 Speech perception in infants. *Science* **171**, 303-318
- Encrevé P. 1988 La liaison avec et sans Enchaînement Paris: Editions du Seuil
- Goldsmith J.A. 1990 Autosegmental and Metrical Phonology Oxford: Blackwell
- Harris J. 1983 Syllable Structure and Stress in Spanish Cambridge, MA: MIT
- Harris J. 1994 English Sound Structure Oxford: Blackwell
- Harris J. & Lindsey G. 1993 The elements of phonological representation. In *New Frontiers in Phonology* (J.Durand & F.Katamba, editors) Harlow, Essex: Longman
- Hyman L.M. 1975 Phonology: Theory and Analysis. New York: Holt, Rinehart & Winston
- 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. & Derrah C. 1987 Representations of speech sounds by young infants. *Developmental Psychology* **23**, 648-654
- Kaye J. & Lowenstamm J. 1985 Compensatory lengthening in Tiberian Hebrew. In *Studies in Compensatory Lengthening* (L.Wetzels and E.Sezer, editors) Dordrecht: Foris
- 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
- 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. 1985 Categorization of speech by infants. In *Neonate Cognition; Beyond the blooming, buzzing confusion* (J.Mehler and R.Fox, editors) 231-262 Hillsdale, NJ: Erlbaum
- 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. 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., Williams K.A., Lacerda F., Stevens, K.N., 7 Lindblom B. 1992 Linguistic experience alters phonetic perception in infants by six months of age. *Science* **255**, 606-608
- Levin J. 1985 A Metrical Theory of Syllabicity (PhD. dissertation) Cambridge, MA: MIT
- McCarthy J. 1979 Formal Problems in Semitic Phonology and Morphology (PhD. dissertation) Cambridge, MA: MIT
- Marr D. 1982 Vision: a Computational Investigation into the Human Representation and Processing of Visual Information San Francisco: W.H.Freeman
- 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
- Otake T., Hatano G., Cutler A., & Mehler J. 1993 Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* **32**, 258-278

Schane S. 1984 The fundamentals of particle phonology. Phonology Yearbook 1, 129-155

- 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. In UCL Working Papers in Linguistics 5 London: University College
- Werker J. & Lalonde C.E. 1988 Cross-language speech perception: initial capabilities and developmental change. *Developmental Psychology* 24, 672-683