# Acoustic Patterns and Speech Acquisition

A.J.Fourcin

Published in Speech and Hearing; Work in Progress, 1978, pages 144 – 171,
University College London, Department of Phonetics & Linguistics *****

By far the greater part of the work at present in progress in the field of speech acquisition depends on two related descriptive tools. The first comes directly from classical phonetics and makes use of place, manner and voice descriptors and a tradition al transcription. These investigations attempt to define the sound contrasts of speech qualitatively, both in production and perception, in what are primarily productive, articulatory, terms. The second method of description uses a particular set of distinctive features (Chomsky and Halle, 1968) which are based on sub sets of these phonetic, articulatory, dimensions. These distinctive features are intended to facilitate the definition of phonological contrasts. This contribution is concerned with a complementary description of some of the aspects of speech acquisition in strictly quantifiable acoustic terms. The acoustic form of speech can be given a direct auditory as well as an articulatory interpretation and this makes it possible to arrive at a realistic appreciation of what elements in a speech sound sequence are likely to be dominant in sensory terms and how these elements must be processed - in normalization for example when listening to a small as opposed to a large vocal tract - so that physically different acoustic stimuli can have a common phonetic identity.

The use of a phonetic transcription necessarily limits the adult investigator and may lead him to assign importance to aspects of a child's speech which are of little contrastive significance to the child himself. The use of quantitative acoustic-auditory descriptors is beginning to reveal aspects of both productive and perceptual processing which could not otherwise have been guessed at. A first example of this, below, is drawn from a study of baby cries (see figure 1). This is followed by a discussion of normalization (see figures 3, 4 and 5). Normalization depends on an ability to perceive similarity of structure - or pattern - and a general indication of the way in which pattern perception may contribute to speech development is given in the discussion relating to figures 6, 7 and 8. The stimuli and data of figures 9, 10 and 11 relate to a particular acoustic study of the way in which English and French children develop their ability to perceive elements of what is phonetically described as the voiced-voiceless distinction. Acoustic patterns not only provide a means for describing speech events but also for the assessment of auditory dysfunction, using synthetic speech, and the correction of inadequate production using pattern displays. This work depends on the possibility of referring to normal acquisition and this is briefly discussed finally.

*****also see: Waterson, N. & Snow, C.E. 1978, pp 47-72
The Development of Communication, John Wiley & Sons, New York

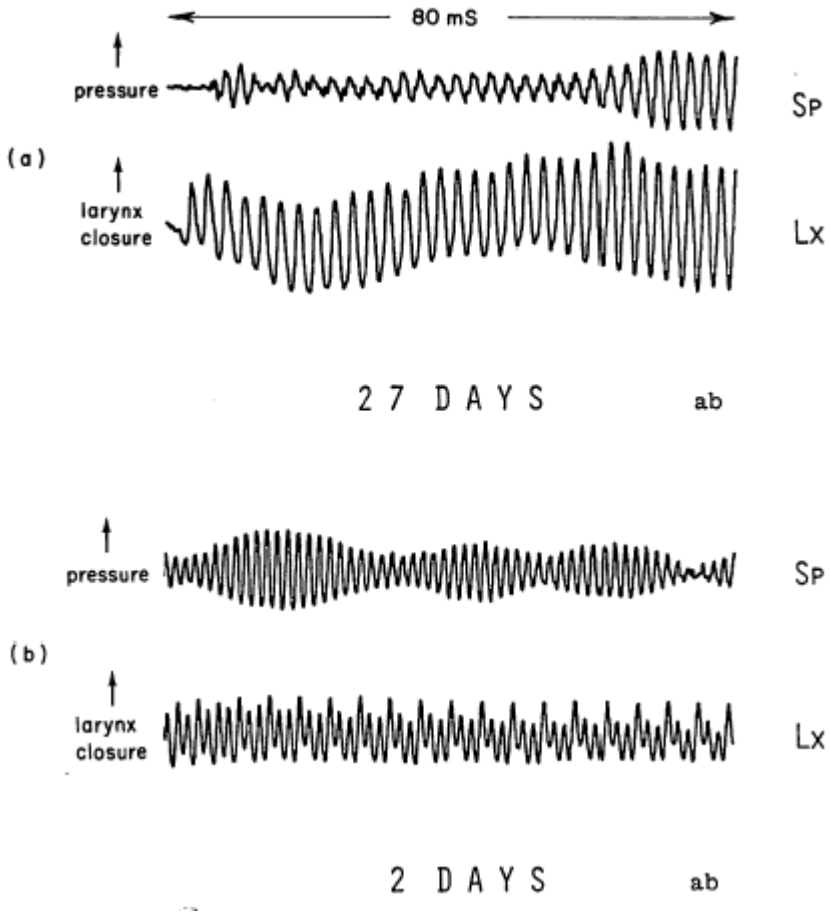


**Fig. 1. Cry development**

*The two pairs of waveforms shown have been taken from a developmental study of one child during the first six weeks after birth (I am grateful to Anne-Britte Parker for making the recordings and to Simona Bennett for her co-operation). In each pair of traces, Sp refers to the acoustic pressure and Lx indicates the output of an electro-laryngograph, simultaneously recorded on an ordinary two channel tape recorder. Laryngeal vibration in the first weeks after birth is not always well defined, and this is shown by the irregularity of Lx in l(b). The Sp waveform shape, however, is determined primarily by the first formant frequency and the relative inadequacy of larynx excitation is responsible only for a small amplitude. When the child has increased his voicing skill and his cry has greater amplitude, his larynx vibration is necessarily more regular and his vocal tract movement necessarily more precise. This is shown clearly in the onset waveforms of l(a). It is important to note that the application of the simple auditory feedback criterion of loudness can guide quite complex productive skills.*

## CRY DEVELOPMENT

The waveforms in the top half of Figure 1 have been recorded from the cry of a 27-day-old baby. Sp refers to the acoustic pressure waveform and Lx designates the synchronously recorded output of an electro-laryngograph (Fourcin, 1974). During normal voicing in both adult and child the vocal folds vibrate regularly, successive closures occur with a quite well-defined periodicity and the maximum glottal opening and greatest degree of vocal fold contact during closure typically vary little from cycle to cycle. The Lx waveform shows this clearly since it is determined primarily by the nature of vocal fold contact

during closure and it can be seen in Figure l(a) that this 27-day-old baby has the type of closure sequence which, in its regularity, corresponds to normal vibration. The frequency of vocal fold vibration , Fx, is markedly higher than that normally found for child and adult (see Figure 3) and starts in this example at about 400 Hz, falling to 340 Hz. Fx determines the fundamental frequency of voiced sounds and is the primary physical correlate of their pitch. The Sp waveform also has some mature features. This can best be seen when the two waveforms in Figure l(a) are interpreted jointly. The Lx waveform starts before there is an appreciable Sp pressure.
This results from the baby's breathstream initiating vocal fold vibration before the release of a vocal tract articulatory closure. Prior to this release both nasal and oral branches of the vocal tract have been held closed, and a controlled oral release has then taken place relatively slowly during the 60 ms interval following the initiation of vocal fold vibration. This sequence of combined laryngeal control and vocal tract gestures is typical in general form - although not in detail - of an initial voiced plosive consonant-vowel combination; it is an essential basis for later contrastive speech productive ability.

The pair of acoustic pressure and vocal fold closure waveforms shown in Figure l(b) have been recorded from the same baby at the age of 2 days. The Sp waveform has a dominant frequency of about 690 Hz and a smoothly fluctuating amplitude which varies as a result of the baby's uncoordinated control of his vocal tract shape. These rapid vocal tract changes - the first two peaks are separated by 30 ms, the second pair by 18 ms - make it very difficult to interpret the formant patterns of the corresponding spectrograms and add to the obstacles which are ordinarily in the way of a spectrographic interpretation of vocal fold excitation. The synchronously recorded Lx waveform is easy to interpret, however. It shows a vocal fold vibration which is quite atypical of the normal mature form. In the adult this regularly repeated sequence of doublets or triplets of decreasing amplitude of closure is found in some of the samples of phonation for unilateral vocal fold palsy (Fourcin, 1974). When the folds are asymmetrically tensed their natural frequencies of vibration may be quite different and they will not act in unison. This can result in a vocal fold version of acoustic beats. A sequence of vocal fold beats will be reset by the relatively violent closure which occurs when the phasing of the folds returns to that of normal vibration. Normally phased vocal fold vibration occurs when the two folds have symmetrical movements; this puts all the acoustic energy into the basic harmonic spectrum and has a greater sound producing efficiency than that of irregular vibration. The Lx waveform of Figure 1(b) in consequence indicates an asymmetric tensioning of the baby's vocal folds which will be associated with a weak cry of ill-defined voice pitch. The triplet sequences of closure which are shown here have a frequency of about 230 Hz whilst the intrinsic vocal fold frequency is about 690 Hz. This difference, if it is

substantiated by other work, could account for the paradoxical developmental increase in pitch of the neonate cry which has been observed to occur for some babies in the first month after birth and partially explain the relative weakness of the cry in this period. An increase in regularity of vocal fold vibration improves the pitch definition and loudness of the cry and both of these features are, in principle, readily capable of mediation by the baby's hearing mechanism. In this case, loudness and pitch are directly related. He can, in consequence, use loudness as a feedback control which will improve his cry in quite detailed aspects of its laryngeal excitation. The uncoordinated control of his vocal tract will also reduce the signalling effectiveness of his cry and can similarly be improved by attention only to the auditory feature of loudness. This factor of auditory feedback must also be of primary consequence in the development of the sound productive skills shown in Figure 2.

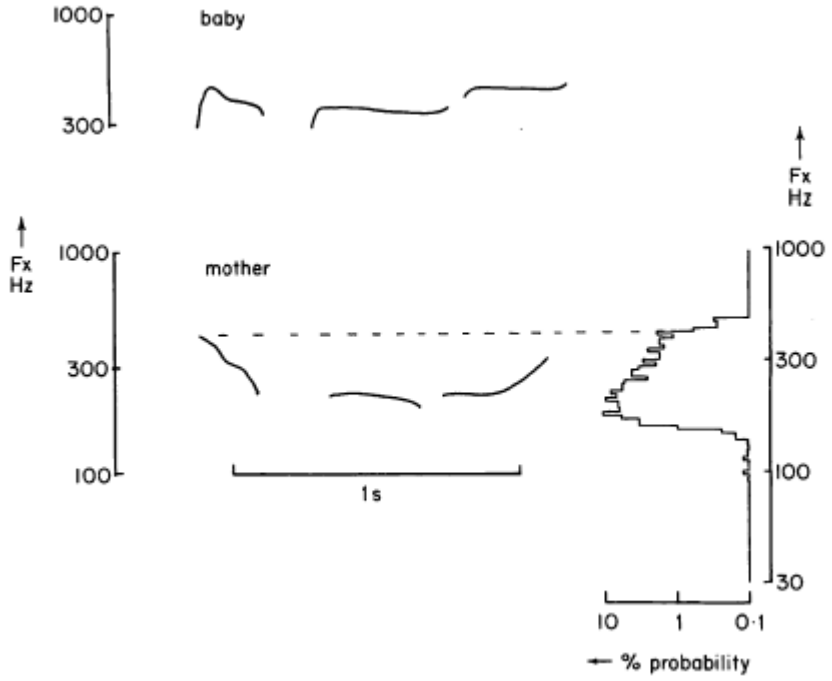


**Fig. 2. Voice pitch interaction**

*The top part of this figure shows the voice frequency contours, Fx, of a particular sequence produced by a 4 week old baby in the company of its mother. Immediately below these three tones are the three voiced segments ([a]) produced by the mother in response to her child. The mother has repeated her baby's sequence with constraints coming partly from the phonology of English and partly, perhaps, from her desire to tune the physical nature of her voice to that of the baby: her fall+rise sequence is a typical English intonation form but is here displaced into an atypical high pitch range. The distribution on the right hand side of the figure shows the range, and probabilities of occurrence, of the voice frequencies in the mother's expressive speaking voice. Her first fall, in this example, starts at a frequency which is at the top extreme of her range. In normal speech this high to mid fall would not occur. Its production here enables her to reduce the complexity of the baby's matching task.*

## IMPORTANCE OF LOW FREQUENCY ENERGY

The essential factor which distinguishes speech sounds from all others which may be produced by the vocal tract, is that they are used contrastively. The basis of contrast is provided by pattern difference and Figure 2 gives an example of the first type of sound pattern which is used by a baby in a controlled way. A sequence of a falling tone, level tone and slightly rising tone is produced by the baby. This is reinforced by the mother and immediately repeated by the infant. In order for the baby of Figure 2 to respond to his mother's utterances and to repeat his own he must be able to make use of at least some aspects of the pitch variations both of her voice and of his own. There is strong empirical evidence that pitch is mediated in the human adult as the result of two distinct types of acoustic signal processing. First and more classically in terms of the place theory, by the positions along the length of the basilar membrane of regions of maximum movement (Newby, 1972). Second, by the transmission along the eighth nerve of time structure information about the acoustic stimulus. When, like the majority of voiced vowels, the acoustic stimulation is a complex waveform with a well-marked period then the frequencies of the fundamental and its harmonics will operate the first pitch mediating mechanism and the periodic waveform irregularities will contribute to triggering the second (Fourcin, 1970). The new-born child has a nearly adult size tympanum and a well formed cochlea (Northern and Downs, 1974). Although a considerable amount of growth dependent development remains to be accomplished, once the middle ear is fluid-free some mechanical cochlear response to acoustic stimulation is to be expected at least at the lower end of the frequency spectrum, since the acoustic impedance match of the immature ear to air may improve as frequency diminishes.

Weir (1976) has examined the results of direct experimental assessments of the auditory frequency sensitivity of the neonate. Her analysis gives credence to the earlier conclusions that stimulation frequencies below 500 Hz and square rather than sine waveforms are most effective in provoking startles in neonates. Although this practical demonstration of the relative effectiveness of low frequency, temporally well defined, acoustic stimulation requires further experimental support; three other factors make it seem possible that the low frequency end of the acoustic spectrum is most important not only to the neonate but also to the young child. The first of these additional factors comes from the preferential masking of high frequency energy by low in hearing; this is a classic result using pure tone stimuli (Wegel and Lane, 1924) and occurs also with voiced formants (Nye, Nearey and Rand, 1974); it appears to result partly from the hydromechanical response of the cochlear partition and is likely to occur in the neonate cochlea as well as in the adult. The second of these factors arises from- a hypothesis (Salus and Salus, 1974) concerning the child's neurophysiological development. The process of myelination is known to

influence the high frequency transmission characteristics of nerve fibres and if myelination is incomplete this might (although not according to a strict place theory of hearing) reduce hearing ability for higher frequencies. Finally, and most certainly, the nature of voiced speech sounds is such that ordinarily there is always greater energy at the fundamental frequency than elsewhere and the first formant ordinarily is greater in amplitude than all others. This physical spectral bias would in consequence act to direct auditory attention to these components of speech.

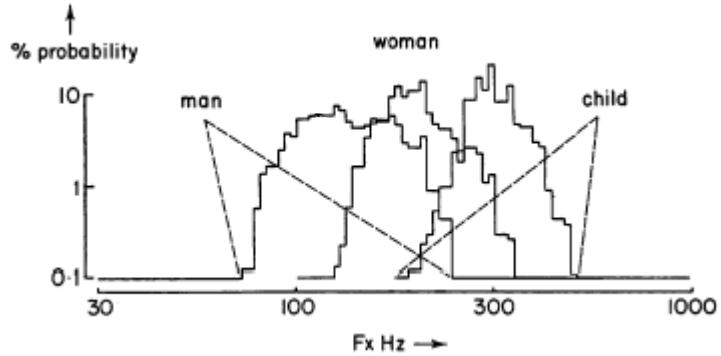


**Fig. 3. Environmental voice frequency ranges**
*Three superimposed voice frequency, Fx, distributions are shown. They have been obtained from three-minute samples of laryngograph waveforms separately recorded by normal man, woman and child speakers. Each speaker produces his or her intonation forms within these physical confines and the developing child must learn to recognise Fx patterns as being identical although their absolute ranges may, as here, be markedly different.*

**FIRST INTERACTIVE COMMUNICATION**

Figure 3 illustrates the quantitative nature of the baby's task when he interacts with the other members of his family solely on the basis of voice pitch. In order to produce the same pattern of change as his father when the father produces a simple falling [a], or to be reinforced when the father imitates the baby's [a], the baby must, as in Figure 2, be capable of pattern rather than absolute imitation. This imitative interaction with the father is likely to be facilitated by a previous extension of the interaction with the baby's mother. Since her normal range of larynx frequencies is already considerably below that of the baby, any successful use of ordinary voice by the mother, in responding to or in eliciting a corresponding response from the baby will contribute to the baby's ability to abstract pattern form. In this way, simple intonation forms produced by parents or siblings can be treated as being perceptually the same and the first step can be made towards the solution of the general problem of acoustic pattern identification. This perceptual congruence is the basis of phonetic identity and it involves a hypothetical processing level which is often referred to as normalization (Fourcin, 1971; 1975).

For the child who is born in a tone language environment we must expect that ease of pitch pattern normalization will provide a first introduction to phonological contrast since contrasting fundamental frequency patterns can be used lexically. At first the contrasts are likely to be crude and oppositions which are the least pitch confusable will precede those which have similar levels and contours.

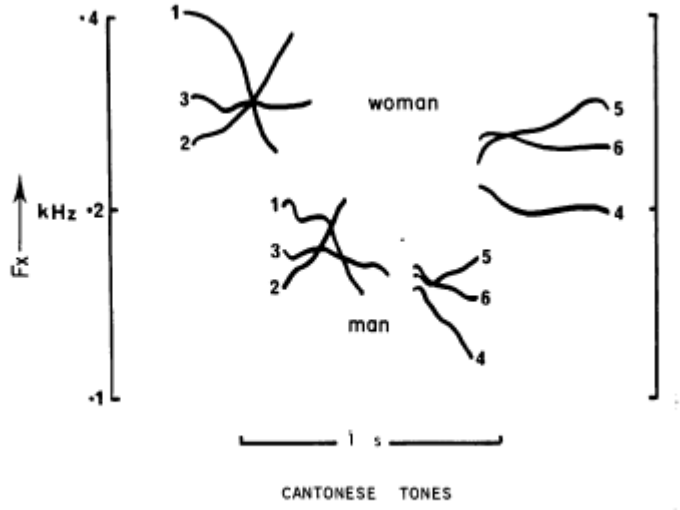


**Fig. 4. Cantonese tones**
*In terms of fundamental frequency structuring, Cantonese has six main tones. The choice of tone by a speaker determines the lexical value of a word. The main pattern relations between the tones - in a given accent - are fixed from one speaker to another and the developing child must learn these relations, and ignore absolute physical differences, in order to perceive and produce lexical tonal contrasts. In the tonal language environment this normalization will be basic to a child's first phonological skills.*

The fundamental frequency basis of lexical tone contrast is shown in Figure 4 for two Cantonese speakers. If, in a tone language environment, the baby's mother restricts her use of articulatory contrast and relies only on the simple pitch distinction which can be based on these Fx contours, we can expect that the first stages of phonetic discrimination will be easier than if the spectral envelope contrasts of non-tonal languages are used. English babies in their first year can make communicative use of voice pitch changes (Ricks, 1975; Lewis, 1968) and it has been commonly observed that at the babbling stage the English child uses English pitch forms. Tone does not have a simple lexical significance in English, however, and the first lexical contrasts depend not on the excitation of the vocal tract but on the spectral envelope of its output. The normalization process must now make use of more complex physical information.

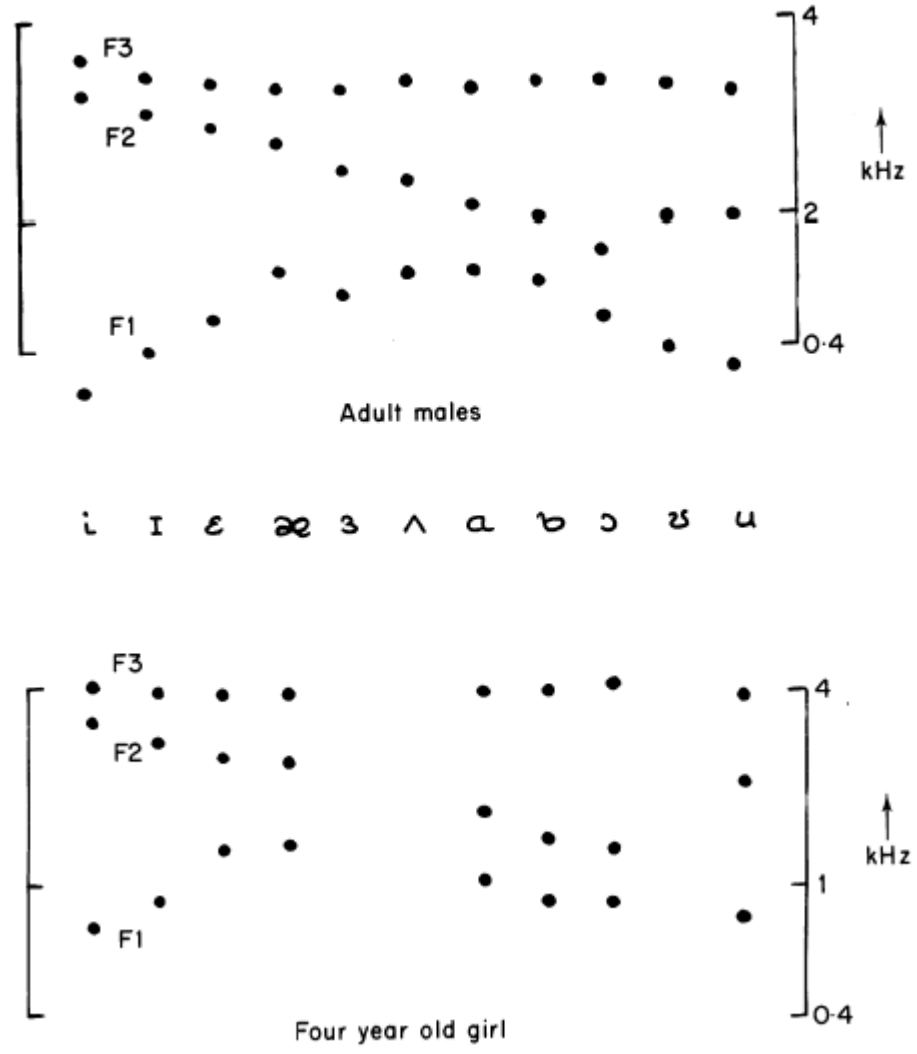


**Fig. 5. English vowel formant frequencies**
*In English, phonological oppositions are carried by vocal tract rather than vocal fold features and tonal differences are of minor lexical importance. Formant frequencies provide the primary acoustic information which enables an auditory assessment of these vocal tract differences to be made and, once more, the listener must allow for physical differences between speech sources in his appraisal of pattern forms. Vowel formant patterns are simpler than consonantal and their essential independence of source is illustrated here.*

Figure 5 shows the average formant frequencies of English vowels produced by young English adult males and, below, the particular values for a four-year-old child. Just as for Cantonese tones, the overall patterning for the two phonetic sets of contrasts is the same although the physical size of the speech sources is markedly different. Although it is generally agreed (Anthony and Mclsaacs, 1970; Sheridan, 1948; Fry, 1966) that the English vowel system is fully acquired long before the consonant system, little is known concerning the pattern of confusions which arises in the early stages of acquisition. In terms only of the first spectral peak, Fl, [i] and [u] are most distinct from [ æ ɜ, ʌ ] and it seems

probable that in the very early stages on the basis of purely auditory information [i] will be highly confusable with [u], as is the case for the deaf child with little high frequency hearing (Fourcin, 1976). Contrasts due to nasalisation, which is associated with primarily low frequency spectral features, will not present an especial perceptual difficulty in early development. Increasing skill in the interpretation of the acoustic signal will enable the position of F2 to disambiguate Fl information for all the vowels. The diphthongs, which are characterized by relatively slow spectral changes, will also be differentiated by this extra spectral information.

Consonant contrasts are all carried by a combination of spectral and relatively rapid temporal differences. The shorter duration of their distinctive elements introduces a variety of difficulties. First, they are more easily masked by external acoustic events since their transient nature reduces the redundancy which is associated with repetition. Second, in the nature of speech production, variability from utterance to utterance, even for a single speaker, is unavoidable and this makes the individual token less well-defined. Third, the sensory processing is handicapped by additional masking, both forward and backward (Elliot, 1971), in time. For example, the initial burst in a voiceless plosive-vowel combination could, in forward masking, reduce the perceptibility of the F2 transition; and in backward masking, the same transition could be masked by the relatively high voice energy in the F2 of the vocalic part. Fourth, the nature of the transitions which characterize consonants will necessarily vary as a function of their context so that their defining patterns and the normalizing processes which are necessary for their retrieval are inevitably more complex than is the case for tone and vowel distinctions. Little has been done to elucidate the perceptual mechanisms which operate at this crucial stage of speech processing (Verbrugge et al., 1976) but the experiments which have been performed (eg Fourcin, 1968) show very large changes in the interpretation by child and adult subjects of identical consonantal stimuli purely as a function of the subjects' inference of the characteristics of the speech sound source. When a child produces a phonetically acceptable consonant-vowel combination he is necessarily using normalization processing either in order to monitor his output or to set up the original reference from adult models. His processing may not, however, be as complete as that employed by a competent adult. At first, the needs of a limited set of phonological oppositions may be served only by attention to Fl and nasal formant transitions. At a later stage, as greater auditory skill is acquired, both Fl and F2 pattern elements could be used and, subsequently, F3 and the fricative formant transitions could be employed in perception and normalization to provide the basis for an essentially complete speech sound inventory. These acoustic auditory pattern considerations do not explicitly include the articulatory constraints which determine ease of production and govern coarticulation and assimilation but their examination in

isolation reveals an aspect of speech development which may prove to be of equal consequence. The child who cannot perceive the relations between the acoustic pattern elements of speech is shut out from ordinary communication.

**NORMAL CONSONANT DEVELOPMENT**

The way in which speech productive skill is acquired by the normally communicating child in an English speaking environment has been studied by a number of investigators both in Britain (for example: Sheridan, 1948; Morley, 1957; Fry, 1966; Waterson, 1978; Anthony, Bogle, Ingrain and Mclsaacs, 1971) and in the U.S.A. (for example: Templin, 1957; Poole, 1934; Wellman, 1931; Menuyk, 1972). A classic phonetic transcription of the material has been employed in all cases and the adult investigators have, of course, applied adult criteria in their categorization of the children's utterances. Although little has been reported in respect of vowel development, each study has yielded results with regard to consonant acquisition and overall there is a useful, and interesting, consensus of opinion.

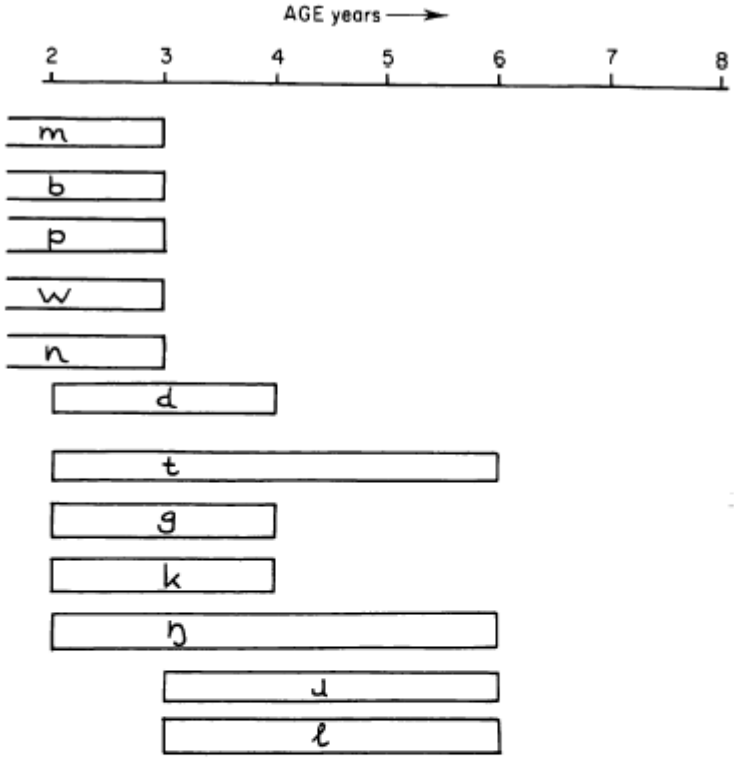


**Fig. 6. Consonant acquisition**

*This summary of English consonant developmental studies is based on a convenient representation introduced by Sander (1972). The left-hand bar for each closed box corresponds to the age at which 50% of the children studied use the sound (ideally this should be a contrastive use); the right hand bar corresponds to the 90% age. Initial, medial and final position occurrences have been averaged, /h/ has been omitted; grouping follows phonetic class.*

Sander (1972) has summarized some of the American work (Wellman and Templin, 1931) and his graphical representation is basic to Figure 6. All workers both in Britain and the U.S.A. find that the voiced and voiceless fricatives and affricates occur towards the end of development and are ahead only of cluster production. [h] is an exception to this rule; in the U.S.A. this is found to occur very early but in the U.K. it is amongst the last observed. The later stages have been omitted from figure 6 since the ordering of acquisition within the fricative class as a whole is not well defined - at least in published reports. The plosives and nasal continuants not only precede the fricatives but also have a fairly well-defined order within themselves. Labials tend to occur before alveolars and these tend to be used contrastively before velars. /w/ occurs with the labials but /I/ and /r/ follow the velars. Only an incomplete definition and only a partial understanding of the factors which lead to this developmental ordering are available at present. Four obvious sources of influence are: speech sound environment; use in communication; ease of production; ease of perception. To an important extent these four appear to fall into two pairs, since the use of speech in the child's environment will be directed towards communication with him and we can expect that early sounds must be readily produced and perceived.

The pressure of sound environment for an English speaking family arises partly from the mere frequency of occurrence of sound types and their contrastive use. The probability of occurrence of sounds in English (Denes, 1963) has been combined in Figure 7 with their median age appearance, using the data on which Figure 6 is based. [t], [n], [d] and [s] are by far the most frequently occurring sounds and their minimal pairs (contrasts such as day-say) also occur most often. It is significant that these alveolar contrasts are so much more frequent in English speech since, even if subsequent work shows that they are not so common in the environment of baby and young child, this result will indicate an important modification of normal speaking habit. The sounds [m], [w], [b], [p] which occur so early are of far lower frequency of occurrence. This fact of early acquisition is not explained in terms of normal environmental pressure on communicative convenience. There is no simple correspondence between the acquisition orders of Figure 6 and the occurrence probability of Figure 7.

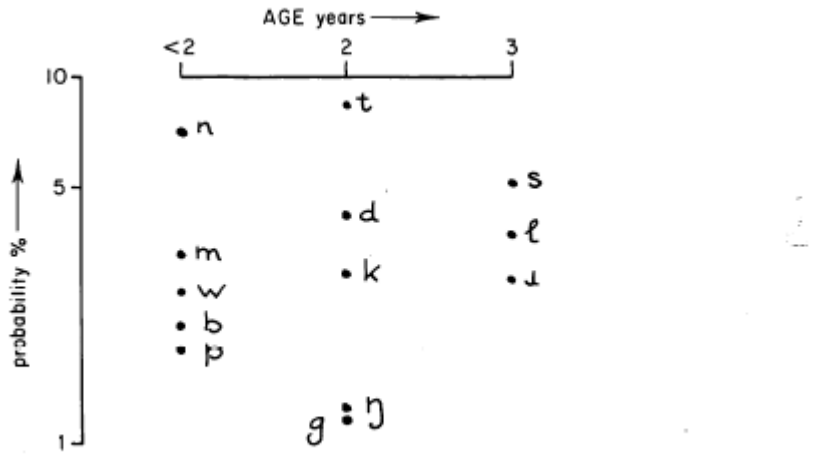


**Fig. 7. Probability of consonant occurrence in the normal speaking environment, against age of (50%) acquisition**
/t/, /n/, /s/, /d/ *occur most frequently in adult speech and provide the most common minimal contrasts (Denes, 1963). This functional pressure does not appear to influence age of acquisition since it can be seen that these and other probable occurrences in developed speech are not necessarily amongst those which are earliest acquired. At present it seems more likely that relative ease of perceptibility is more important in early development than phonological pressure.*

In early production consonants occur most frequently in initial rather than in final position. For the voiced plosive consonants this requires a moderate degree of coordination between laryngeal and vocal tract controls to be exerted by the speaker. The speaker's soft palate must be raised so that an oral pressure increase can be established, the vocal folds approximated to their position for free vibration and then the airstream can be initiated. No fine adjustment is needed and the baby's early sucking and crying abilities are directly applicable to this speech skill. For the production of [m] the control sequence is simpler since the closure is maintained instead of being released rapidly as for [b]. [w] is obtained by using the controls for [b] but associating them with a much slower movement and an incomplete vocal tract closure. These bilabials are the simplest consonants in productive terms and their simplicity may well have a bearing on their early appearance in the young child's speech. An important difficulty arises in the case of the voiceless bilabial [p], which requires a much greater degree of productive skill in the simultaneous control of vocal tract and vocal folds so that oral pressure is built up and released before the onset of vocal fold vibration (Stevens, 1971 ; Kewley-Port and Preston, 1974). In spite of this considerable additional complexity [p] occurs before the productively simpler [d]. When the potential ease of perception of these different consonants' acoustic patterns is considered a complementary explanation is found which goes some way towards resolving this dilemma. For initial [b] the burst of acoustic energy which accompanies the release of articulatory closure occurs

essentially together with the voicing excitation of the formants. Fl, as for all initial voiced plosive consonants, starts from a low value and increases quite rapidly to its value for the accompanying vowel. This pattern of change in Fl is a primary acoustic trait for this consonant class. F2 also increases to the vowel value, from a lower frequency which, when taken in conjunction with the set of Fl and F2 frequencies characteristic of the particular speaker, can be found from an inferable locus - a concept given its first quantitative definition in work at the Haskins Laboratories (Delattre, et al., 1955). This pattern of change in F2 can be regarded as a secondary trait, of relevance only for distinguishing [b] from [d] and [g]. If a young child has a greater facility for the processing of low rather than high frequency information in a stimulus having several formant peaks, then he is more likely to be able to perceive formant patterns which have F2 in the low end of the frequency spectrum than in the high.

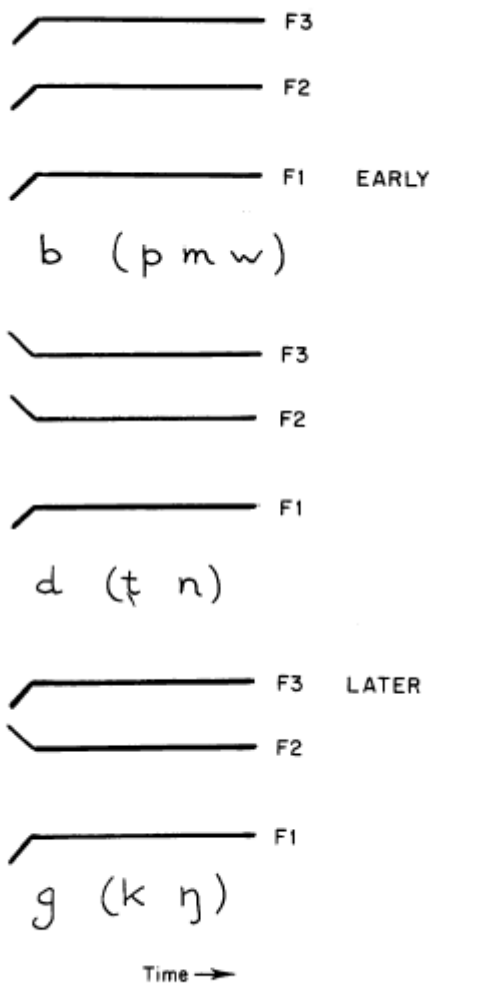


**Fig. 8. Schematic Formant-Time Patterns**
*The early development of bilabial consonants may be influenced both by visibility and auditory clarity. Simplicity of consonant acoustic patterning in general, however, appears to be related to ease of acquisition. It may prove to be of considerable significance that the alveolar and velar patterns are acquired later, not merely because of their relative lack of visual cues but also because of their relative acoustic pattern complexity. (These consonant patterns could be associated with a front open vowel produced by the child of Fig. 5.)*

In Figure 8 the formant patterning for initial [b] is drawn in a highly schematized way but it is clear that the F2 for [b] will be less masked than that for either [d] or [g]. This greater sensory clarity of the [b] pattern may be important in the early stages of speech perception in enabling the contrasts between [b], [m] and [w] to be established primarily on the basis of Fl changes in time and secondarily with the aid of F2 as a source of reinforcement. Production of these sounds will then be facilitated by the auditory feedback made possible by a simple set of pattern references which require little cognitive elaboration for their successful application to a wide range of speech sources of different vocal tract dimensions. In this way, at this first level (Fourcin, 1971; Lisker and Abramson, 1964) and first developmental stage of speech processing, normalization can be established as the joint result of perception and associated production. For initial [d] the formant patterning in Figure 8 shows a falling transition for both F2 and F3 and, as greater auditory skill is acquired this reinforcing F2-F3 alveolar fall may be contrasted with the reinforcing bilabial rise in F2-F3 which typifies [b]. This is a much simpler opposition in acoustic pattern terms than that between [b] and [g] or [d] and [g], since F2 and F3 for velars tend to move in opposite directions. From a purely auditory-acoustic pattern point of view this makes it quite likely that [d] and its associated nasal [n] will be next employed contrastively with each other and with the previously acquired bilabials. For initial [p], although the skilled adult may make use of more than a dozen different acoustic traits, the very young child is likely only to be influenced by the most evident pattern change. Following the release burst for [p] there is typically an interval in which no voicing occurs and only aspiration excites the speaker's vocal tract. The aspiration gives relatively less energy to Fl than to F2 and F3 compared with vocal fold excitation, and the F2, and possibly the reinforcing F3, transitions may be utilisable by the child as a secondary trait. The gross trait of lack of initial voicing dominates for [p], [t] and [k] and is a primary characteristic. At the beginning of speech acquisition when attention is directed essentially to the Fl region the voicing gap provides a simple way of including [p] in the family of contrasting bilabials. With increasing auditory skill the secondary F2-F3 information can be utilized and, depending on individual circumstance and vowel environment, the difference in burst frequency, which exists between bilabial and alveolar initial plosives, may be utilized. In acoustic pattern terms initial [k] is most confusable with initial [t] (and this confusion will be greatest for a high F2 front close vowel environment). With the increasing auditory skill, which comes with the gradual approach to speech maturity, however, the more complex F2-F3 patterns from the velars will be re solved and the family of velar contrasts will be added more consistently to the alveolars and bilabials. This brief discussion of the possible relevance of acoustic pattern forms to the ordering of speech sound acquisition has concentrated on only the most obvious facts. It is evident, however, that the correspondence between the

ordering implied by the pattern sequence of Figure 8 and that of Figure 6 is far greater than that implied by the occurrence probability structure of Figure 7. It is not possible using either the standard techniques of transcriptive phonetic analysis or distinctive feature categories (Chomsky and Halle, 1968) to examine the way in which children's speech development is influenced by these acoustic pattern forms. It is feasible, however, to use processes of acoustic analysis to measure the evolution of productive skill and to use speech synthesis techniques to assess children's ability to perceive acoustic pattern differences which have speech significance. Both of these methods have been employed in this department (University College London) for normal children (Simon, Simon and Fourcin) and for the deaf child. The first results of this work indicate that the stages of speech development and the influence of auditory disability are better understood when acoustic pattern description is allied with phonetic analysis. The phonetician necessarily applies an analysis which reflects his ability to hear whole pattern forms; he is intrinsically unable as the result of his training to attend to the pattern elements which may dominate a child's perception of particular contrasts and which give rise to such phonetically strange contrasts in the young child's speech repertoire. The following example is concerned with what phonetically is termed the 'voiced-voiceless contrast.

## PATTERN PERCEPTUAL DEVELOPMENT

The acoustic pattern forms corresponding to a particular example of this voiced-voiceless contrast are shown in Figure 9. The voiceless extreme (V-) is on the right, the voiced (V+) is at the lower left. Acoustically, a large number of factors underlie this simple phonetic opposition (e.g. for V+, an initial upward step in Fx; a rising Fl, a lower intensity burst which occurs close to the onset of voicing; and for V- an initial turbulent excitation following the burst which is often described as a delay in voicing, an initial downward fall in Fx with an initial breathy excitation, an effective initial absence of Fl and an initially greater burst intensity). Other factors can be listed particularly as a function of vocalic environment but only two of all those possible are explicitly dealt with in the figure. The first arises from the delay in voicing - this is called voice onset time, VOT (Lisker and Abramson, 1964). The second arises from the normally rising Fl for V+. In the top part of the figure a flat Fl is shown for V+; this is not a naturally produceable pattern form but for the deaf listener with high frequency loss it can be used to elicit a V- response (Simon, Simon and Fourcin). This type of listener cannot respond to patterning associated with the burst and can only contrast Fl pattern forms.

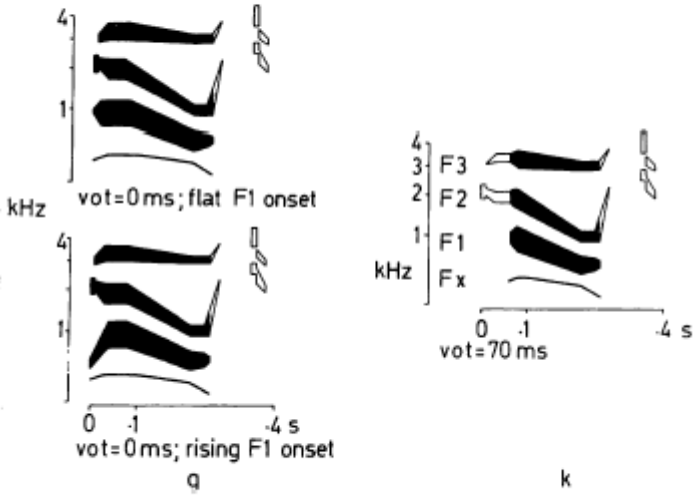


**Fig. 9. Synthetic pattern extremes for : [g] and [k]**

*These patterns are particular examples of stimuli which can be used in a controlled way to examine a listener's ability to make contrastive use of particular components in speech. The left hand patterns represent 'goat' , the right-hand pattern 'coat' . The flat Fl onset stimuli cannot be produced naturally but it proves to be an acoustic pattern feature which can be employed perceptually in speech sound discrimination by the deaf - to infer lack of initial voicing;* [k] *labels are then given both to these flat Fl stimuli as well as those of the form shown on the right. In the next two figures, average responses to these rising Fl transition stimuli are shown by squares, the crosses represent responses to flat Fl and large voice onset time delays.*

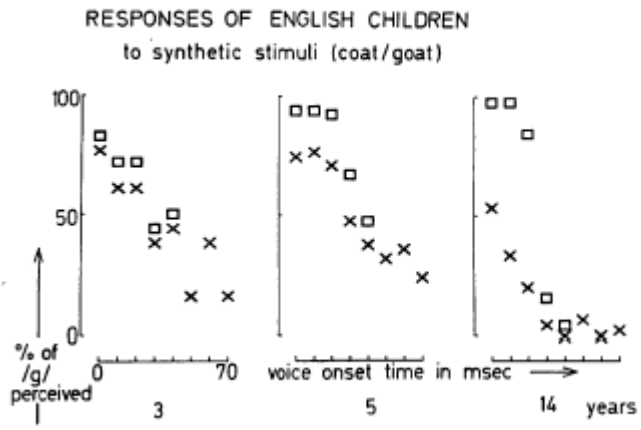


**Fig. 10. Voice-voiceless perception by English children**
*The ability to discriminate between initial* [g] and [k] *is only acquired, on average, gradually. At first the pattern features which are most obvious guide the child's labelling and, for the three year-old children, there is little distinction between formant shapes, only the degree of periodic excitation is of real importance. With increasing age there is increasing skill both in labelling and in the ability to reject non speech-like patterning, the cross stimuli are not, in consequence, often put in the* [g] *category.*

In Figure 10, (Simon & Fourcin) , the rising Fl stimulus responses are shown by squares and for the 14-year old children it can be seen these stimuli evoke a sharp categorical response. The average responses to the flat Fl stimuli in the same VOT range are shown by crosses, these are not well categorized. The 3-year-old children, however, categorize both of these stimulus types in essentially the same way; they are responsive to VOT and do not make any special use of the rising Fl information. This is not the case for the 5 year-old children who are in an intermediate state of development. These results show how two acoustic traits can be used differently as perceptual development proceeds. The learning is gradual and it can be easily interpreted in terms of acoustic pattern salience but not at all in classic phonetic or distinctive feature terms. In consequence a child could make consistent contrasts which are not understandable with normal analytic techniques.

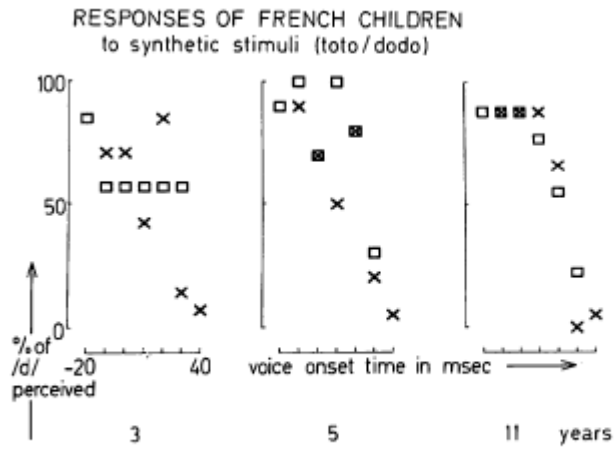


**Fig. 11. Voice-voiceless perception by French children**

*French does not rely on the presence or absence of aspiration, and the associated delay in voicing onset, to provide a basis for the voiced-voiceless distinction. Pre-voicing, before release, characterizes the voiced sounds. This provides a quite different perceptual bias, as compared with English, and the French children of all ages are not markedly influenced by Fl shape but are progressively more skilled in using the onset timing of periodic excitation.*

In Figure 11 (Simon, and Simon & Fourcin), the results are shown from the same type of experiment performed with French children. The speech environment of these children does not employ the post-burst turbulence used in English and the presence of a rising Fl is not a useful contrastive trait. This is reflected in the uniformity of response by the 11-year-old children to both rising and flat Fl stimuli. The same similarity of response is found with the 5-year-old children. It is evident from these two language different experiments that response is influenced by the models afforded by the speaking environment and that categorization skill may be acquired quite slowly. The greatest future value of speech pattern tests may come, however, from the quantitative information which they provide concerning individual ability and development - both normal and handicapped.

## DISCUSSION

The work which has been described has all been concerned with a study of various aspects of speech communication, as opposed to mere discrimination. It appears quite likely that the acoustic auditory description, and definition, of prominent aspects of speech sound combinations will be of real help in understanding the increase in skill which is basic to the developmental process underlying the progression from the first cry to the complete mastery of phonological oppositions.

At each stage of development, prominent acoustic traits will be used by the individual listener both in assessing his own production and that of other speakers. His processing of other source outputs will depend on his ability to normalize, and his ability to use sounds contrastively will depend on his ability to categorize. Rules of acoustic pattern processing are needed for both of these aspects of perception and, in due course, we must be able to formulate grammatical systems for these pattern relations both to understand the process of speech acquisition more fully and to be able more adequately to describe normal adult usage. An outstanding apparent anomaly exists, however. Work, especially by Eimas (1975) has shown that the very young infant can discriminate speech sounds in ways which appear to reflect an innate predisposition to categorization. The discriminations, however, are sensory rather than phonetic.

The work described here shows only a gradual accumulation of speech processing ability. The experiments that Eimas has performed depend on the baby's sensory ability to discriminate between sound patterns presented in sequence. Later on it seems likely, from the present discussion, that a ready ability to normalize on the basis of only a little prior experience will also be found. It does not follow, however, that the ability to categorize speech sounds contrastively, so that communication ability is achieved in a particular language environment, can be innate. This is a higher skill which must be learnt from experience, and which could well in part depend on an innate auditory ability to process prominent acoustic pattern traits.

We must expect, for example, that the categorization of VOT differences will partly depend simply on the peripheral hearing mechanism's response to transient stimulation (a 20-30 ms minimum stimulus duration is required for accurate pitch assessment and this corresponds to the VOT labelling transition). In consequence, elementary VOT discrimination will be possible for all animals having similar cochlear characteristics.
Similarly, the critical band response characteristics of the cochlea will make a major contribution to the sensory evaluation of formant energy concentrations

and, in consequence, partly determine the ability to detect formant transitions, by both man and animal. These are innately determined characteristics. Normalization and phoneme categorization abilities will be acquired from experience, however. This first without difficulty, by animals as well as infants, since only elementary pattern processing is needed. The second with increasing difficulty as the degree of pattern complexity becomes greater (e.g. in going from labials to velars).

A practical application of speech pattern descriptors is beginning to be made in the remediation of speech productive disability (Fourcin, 1974; Abberton and Fourcin, 1975) and in the assessment of hearing for speech (Fourcin, 1976). It seems possible that future work based on the acoustic analysis of speech, in terms which are of auditory significance, will be a major source of knowledge of both normal speech development and of means for its encouragement.

**REFERENCES**

Abberton, E.R.M. and Fourcin, A.J. (1975) Visual Feedback and the Acquisition of Intonation. In E.H. Lenneberg and E. Lenneberg (eds.) Foundations of Language Development. Academic Press, New York.

Anthony, N. and Mclsaacs, M.W. (1970) Notes on Patterns of Development found by using the Qualitative Phonetic Assessment Sheet of the Edinburgh Articulation Test. British Journal of Disorders of Communication 5^. 148-164.

Anthony, N., Bogle, D., Ingram, T.T.S. and Mclsaacs, M.W. (1971) Edinburgh Articulation Test. E.S. Livingstone"; Edinburgh and London.

Chomsky, N. and Halle, M. (1968) The Sound Pattern of English. Harper and Row, New York.

Delattre, P.C., et al. (1955) Acoustic Loci and Transitional Cues for Consonants. Journal of the Acoustical Society of America 27. 769-773.

Denes, P.B. (1963) On the Statistics of Spoken English. Journal of the Acoustical Society of America ZS~. 892-904.

Eimas, P.D. (1975) Speech Perception in Early Infancy. In L.B. Cohen and P. Salapatek (eds.) Infant Perception. Academic Press, New York.

Elliot, L.L. (1971) Audiology Backward and Forward Masking, 10. 65-76.
Fourcin, A.J. (1968) Speech Source Inference. Institute of Electrical and Electronic Engineers. AU-16. 65-67.

Fourcin, A.J. (1970) Central Pitch and Auditory Lateralization. In R. Plomp and G.F. Smoorenburg (eds.) Frequency Analysis and Periodicity Detection in Hearing. A.W. Sijthoff, Leiden, The Netherlands.

Fourcin, A.J. (1971) Perceptual Mechanisms at the First Level of Speech Processing. In A. Rignault (ed.) Proceedings of the VII International Congress of Phonetic Sciences. Mouton, Lattay: Paris.

Fourcin, A.J. (1974) Laryngographic Examination of Vocal Fold Vibration. In B. Wyke (ed.) Ventilatory and Phonatory Control Systems. Oxford University Press, London. Fourcin, A.J. (1975) Speech Perception'in the Absence of Speech Productive Ability. In N. O'Connor (ed.) Language, Cognitive Deficits, and Retardation. Butterworths, London.

Fourcin, A.J. (1976) Speech Pattern Tests for Deaf Children. In S.D.G. Stephens (ed.) Disorders of Auditory Function II. Academic Press, New York.

Fry, D.B. (1966) The Development of the Phonological System in the Normal and Deaf Child. In F. Smith and G.A. Miller (eds.) The Genesis of Language. M.I.T. Press, Camb., Mass.

Kewley-Port, D. and Preston, M.S. (1974) Early Apical Stop Production: Voice Onset Time Analysis. Journal of Phonetics 2. 195-210. Lewis, M.M. (1968) Infant Speech. Paul, London. Routledge and Kegan

Lisker, L. and Abramson, A.S. (1964) A Cross-language Study of Voicing in Initial Stops. Word 20. 384-422.

Menyuk, P. (1972) The Development of Speech. Bobbs-Merril, Studies in Communicative Disorders (Library of Congress, ref. 74- 173981).
Morley, M. (1957) The Development and Disorders of Speech in Childhood. Livingstone, Edinburgh. Newby, H.A. (1972) Audiology. Crofts, New York. Appleton-Century- Northern, J. and Downs, M. (1974) Hearing in Child ren. Williams and Wilkins, Baltimore.

Nye, P.W., Nearey, T.M. and Rand, T.C. (1974) Dichotic Release from Masking: Further Results from Studies with Synthetic Speech Stimuli. Haskins Laboratories. Status Report on Speech Research 37/38 123-138.

Poole, I. (1934) Genetic Development of Articulation of Consonant Sounds in Speech. Elementary English Review :U. 159-161.

Ricks, D. (1975) Vocal Communication in Pre-Verbal Normal and Autistic Children. In N. OfConnor (ed.) Language, Cognitive Deficits and Re tardation" Butterworths, London.

Salus, P.H. and Salus, M.W. (1974) Developmental Neurophysiology and Phonological Acquisition Order. Language 50. 151-160.

Sander, E.K. (1972) When are Speech Sounds Learned? Journal of Speech and Hearing Disorders JT. 55-63.

Sheridan, M.D. (1948) The Child's Hearing for Speech. Methuen, London.

Simon, C. (1976) A Developmental Study of Acoustic Pattern Production and Perception in Voiced - Voiceless Oppositions. Ph.D. thesis-for the University of London.

Simon, C. and Fourcin, A. (1978) Cross-language study of speech-pattern learning. J. Acoust. Soc. Am. Volume 63, Issue 3, pp. 925-935 (1978)

Stevens, K.N. (1971) Aerodynamic and Acoustic Events at the Release of Stop and Fricative Consonants. Journal of the Acoustical Society of America 50 139.

Templin, M. (1957) Certain Language Skills in Children. University of Minnesota Press, Minneapolis.

Verbrugge, R.R. et al. (1976) What Information enables a Listener to Map a Talker's Vowel Space? Journal of the Acoustical Society of America 60. 198-212.

Wegel, R.L. and Lane, C.E. (1924) The Auditory Masking of One Pure Tone by Another and its Probable Relation to the Dynamics of the Inner Ear. Physical Review 23. 266-285.

Weir, C.G. (1976) Auditory Frequency Sensitivity in the Neonate: a Signal Detection Analysis. Journal of Experimental Child Psychology 21. 219-225.

Wellman, B. et al. (1931) Speech Sounds of Young Children. University of Iowa Studies in Child Welfare 5. 1-82.