'FIR' DIGITAL FILTER MODELLING OF LINGUISTIC PHARYNGEALS TRANSMISSION: SHORT-TIME AMPLITUDE IN LOW-FREQUENCY RAPID SPECTRUM CHANGE

Brenda Orser

Department of Linguistics
University of Victoria

1. INTRODUCTION*

This paper is a report on research into the preliminary design and application of FIR (finite impulse response) filters using CSL © (Dickson and Clayards 1990). These are used to quantize the sound pressure level (SPL) amplitude power component of linguistic resonant pharyngeal consonants / R /, in root-initial CV environments, Fig. 1, / R aymt/ 'angry'. The data are from the phonological inventory of Spokane, Interior Salish, recorded by Carlson (1992, 1969). Voiced pharyngeal resonants are limited in their distribution cross-linguistically (Maddieson 1984). Spokane was selected due to the morphophonemic occurrence of pharyngeals in that language, and because within Spokane there is a measureable phonetic, surface difference between early and recent data collected on root-initial pharyngeal segments, Figs. 1 and 2. In both samples there is ambiguity between the pharyngeal resonant, laryngealized (creaky) voice, and the glottal stop, / G /. In earlier recordings the pharyngeal is clearly the primary segment, and laryngealization is negligible. This differs in recent data: the glottal stop is the primary segment, often accompanied by creaky voice, and with no pharyngeal resonant. At issue then is the integrity of the pharyngeal, its status as a segment and its acoustic characteristics whereby it may be identified as a pharyngeal.

The objective in this paper is primarily descriptive: to provide acoustic specifications of the pharyngeal resonants. This is obtained in noting the amplitude response to filters applied to the pharyngeal segment, wherein the low-frequency spectrum is included in the filter, and transients are captured within this range. Second, in observing the acoustic behaviour of the filtered segments, a relationship is analysed between the fundamental, the lower harmonics, and the first formant. This in turn contributes to the definition of the pharyngeal resonant. Further, its contrastiveness with phonetic glottal onsets is hypothesized on a comparative basis. Lastly, the acoustic specifications of the pharyngeal resonant are drawn, interpretively, stating their function informally in terms of their transmission. This interpretation is tentative; but I believe it is well-founded, in the analysis of pharyngeals within a framework of sound transmission.

In the research here, digital filters are used to develop a reliable psychoacoustic model comparable to linguistic distinctive features specific to pharyngeals in root-initial syllable onset position. While this requires an evaluation of psychoacoustic scales presently in use, the evaluation of such is outside the immediate scope of this paper. For present purposes the accepted measurement unit pertaining to amplitude dynamics is the 'Bark', a frequency-based scalar unit from which the 'Bark Scale' is derived. The scale's unit, the Bark, has an acoustic equivalent: 1 Bark = 100Hz. The frequency range of concern in this research is from 200Hz to 950Hz, to include the fundamental frequency, i.e. [voice], and the first and second formants which in addition to LPC frequency spectrum (formant structure) features, can be interpreted in terms of articulatory features.

* The letter R is used to denote the pharyngeal resonant; G denotes the glottal stop.

In applying filters to digitized language segments, the amplitude response of the Fig. 1 onset consonant only is modelled, with FIR transfer functions corresponding to specific SPL acoustic responses: glottal excitation (forced response) specific to phonation (voicing), Fig. 3, which is separate from the pharyngeal transmission (natural response), specific to articulation, Fig. 4. These responses are assumed generally to be present as input, at the microphone.

The FIR transfer function (impulse response) system design has the following properties: pharyngeal transmission interpreted as odd amplitude/even phase, differentiated from glottal source characteristics: even amplitude/odd phase. This motivates the application of a specific digital structure referred to as a DTLTI: discrete time, linear time-invariant system. Its important design characteristic in modelling pharyngeal transmission is that the amplitude is odd (nonlinear) and the phase is even (linear). This design specification is a departure from the literature on psychoacoustic and digital signal processing of speech.

The objective of the FIR design as applied to speech transmission, pharyngeals in particular, is to derive linguistic structure, along general lines, such that transmission functions analysed in the signal correspond to segmentation in the auditory system and to active articulatory mechanisms in speech production. In the case of pharyngeals one general issue is the separation of the system from the source: here, a transmission function for decoupling vocal cavity resonances from vibrations of the vocal cords (Fant 1970, 1980; Fant and Pauli 1974). The basic design problem is the filter order: the determinants of the SPL amplitude constant derived from an impulse which, integral to the pharyngeal transfer function, will satisfy a linguistic system's transmission requirements taken from psychoacoustic literature as a necessary condition in the auditory analysis of pharyngeal resonants. As acoustic transients, resonants exhibit an increased and rapid spectrum change of amplitude in the low-mid frequency region, Fig. 4. Also, decreasing energy in the glottal excitation contrasts acoustically with the increasing energy exhibited in the transient, (Stevens and Keyser 1989; Stevens and Perkell 1977), (Figures not included).

2. FILTERS

Neither the (temporal) rapidity of the spectral change nor the bandwidth are modelled, since the problem is the nonlinearity of SPL amplitude and not frequency characteristics. Instead, the real-time dimension and the bandwidth are derived as constants, which are linear generally. An FIR filter structure is implemented to obtain the linear phase and thus a constant phase shift, thereby allowing the amplitude response to be modelled. In an FIR filter the impulse response $h(n)$ is therefore limited to a finite number of points in real-time (Stanley, Dougherty and Dougherty 1984; including following equations). The impulse response is expressed as

(1) $$h(n) = \begin{cases} \alpha_n, & \text{for } 0 \leq n \leq k; \\ 0, & \text{elsewhere.} \end{cases}$$

or

(2) $$h(n) = \sum_{i=0}^{k} \alpha_i \delta(n-i).$$

The transfer function for (1) or (2) is expressed as

(3) $$H(z) = \sum_{m=0}^{k} \alpha_i m_{m-z} = \alpha_0 + \alpha_1 z^{-1} + \cdots + \alpha_k z^{-k},$$

where k = the number of terms, and function order. The difference equation for (3) relating the output to the input is

(4) $$y(n) = \sum_{i=0}^{k} \alpha_i x(n-i) = \sum_{i=0}^{k} h(i)x(n-i).$$

This describes a nonrecursive realization for the FIR transfer function.

One significant advantage of the FIR filter function is its capability in obtaining linear phase (i.e., constant time delay), derived via expanding the amplitude response in a sine series. The transfer function also is expanded in sine terms; the amplitude response can then assume negative values. This method is possible because at low frequencies the amplitude response is asymptotic to w^k, with k odd (Stanley, et. al. ibid: 223; Lagerstrom 1988). This is a systemic transmission property.

The FIR filter amplitude response approximating that of an ideal differentiator is

(5) $$A(f) = w,$$

(6) $$\beta(f) = \pi/2 - MTw = \pi/2 - 2\pi MTf.$$

where w = radian frequency (π/s); $A(f)$ = amplitude response; and $\beta(f)$ = phase response, in radians: the phase associated with the noncausal function combined with the additional phase due to the added delay (Stanley, et. al. ibid: 218–226). This method differs from stating the filter function in cosine terms, the amplitude response of which has a constant amplitude slope, and thus is predictable for any odd phase passband. Segments bearing this type of acoustic feature, and without phonatory input, are non-contrastive linguistically, and are attributed to vowel colouring, or, secondary features. The segments are contrastive in some languages only if phonatory features are added (Esling, et. al. 1991; Laver 1988).

Acoustic analysis using cosine terms thus does not capture the transient's amplitude change. Also, whereas the phase shift in cosine terms is ideal, the sine-type expansion allows for a real linear phase: its phase shift is constant and at 90°; also as per (6) it is given in radians, etc., and has all its poles at the origin, and therefore is stable.

The FIR filter order determines the transfer function as a real function of frequency, thus proferring stability on the linear phase characteristics. In a DTLTI system the FIR amplitude response, $A(f)$, is related as a transfer function not only to the phase response, $\beta(f)$, also to the impulse response. The latter, approximating a real function of frequency (i.e., critical band), should satisfy the joint requirement for transmission: linear phase and constant time delay. The impulse response is multiplied in a Hamming window function, with a 12th order bandpass filter. A low-high cutoff is used, where 1 = 5000Hz, the Nyquist frequency. The filters are applied to the surface phonetic form: the onset consonant segment, [R]. In one filter the bandpass cutoff is .04–.19 (200–950Hz: 7.5 Bark), (Fig. 5); in a second filter (Fig. 6) the cutoff is .10–.19 (650–950Hz: 3 Bark). Thus in the first filter both the fundamental and the bandwidth approximating the second formant are included; in the latter filter only the second formant bandwith is included. The research question then is, how narrow must the main lobe of the window be so as to satisfy sharp tuning, active cochlear mechanics in the auditory system, whereby short-time amplitude in low-frequency spectrum change is detected and identified? (cf., Pickles 1991, for discussions on sharp tuning).

3. RESULTS and SUMMARY

Both filters when applied to the entire input waveform of the pharyngeal consonant onset segment, including the phonetic glottal onset, as in Fig. 1 result in a null amplitude response and formant/resonance response to the frequency range, Figs. 7 and 8, respectively. Also, the pitch characteristics are the same for both filters, i.e., pitch is the same with and without the fundamental. Their pitch patterns differ, however, from the pre-filtered pitch characteristics in Fig's. 3 and 4, the phonetic glottal onset and the pharyngeal resonant, respectively, in which pitch patterns were identical. Additionally, the subjective perceptual sound of the segment in each instance of filtering is a woody pulse — flatly-tuned and lacking perceptual resonance. In sum, without the fundamental the filtered output of the pharyngeal segment as a whole, including the phonetic glottal onset, is similar acoustically and perceptually to a glottal stop, [G] in which there is minimal laryngeal activity, ideally none — i.e., [-voice], and a constricted glottis [+cg], Fig. 2.

Fig's. 9 and 10 illustrate a similar filter-response, but with each filter applied to the phonetic glottal onset as segmented in Fig. 3. In both applications, with and without the fundamental frequency, there is an insignificant, if any, amplitude response. Fig. 9, illustrating the fundamental included in the filter, shows energy (in D Window), though this is low and a constant function of the entire frequency range, 200–950Hz. The energy response is not necessarily a response to the fundamental; in Fig. 9 to the fundamental's presence, in fig. 10 to its absence. The nil response is conceivably dependent on the presence of the first formant in the low frequency region, i.e., 300-650Hz. In contrast in Fig. 10 the bandpass without the fundamental and the first formant, there is no energy response nor is there an amplitude response. This then suggests that energy in the phonetic glottal onset is related to the presence of both the fundamental and the first formant. Similarly, the nil FFT/SPL amplitude response varies relative to the presence/absence of the fundamental and the first formant.

In comparison, Fig's. 11 and 12 illustrate the response to both filters to the onset and transient pharyngeal component of the segment analysed from Fig. 4. In both applications there is an amplitude response. Though this is comparatively lower when the fundamental and first formant are not included, Fig. 12, there is an amplitude response in both applications. Further, the damping configuration is similar to the overall shape of the pre-filtered pharyngeal, Fig. 4. Lastly, in comparison with the low and constant energy in the phonetic glottal onset filter including the fundamental and first formant, Fig. 7, the pharyngeal's energy function is low but increasing and stepped. However, similar to Fig. 8, the filter not including the fundamental and first formant, has no energy response.

In sum then, the energy response to filters on the phonetic glottal onset and the transient pharyngeal is dependent hypothetically on the fundamental and the first formant. The dependency of the glottal onset's SPL amplitude response on either the fundamental or the formant structure is indeterminate, as is its dependency on the presence or absence of the fundamental alone.

Comparatively, pharyngeal resonants do show an SPL amplitude response, when filtered with and without the fundamental and the first formant. While this does not provide full support for mutual and hierarchical dependency relations, the acoustic observations do narrow the relationship to articulatory terms equivalent to those involved in the function of a relation between the fundamental, first formant, and SPL amplitude response. This suggests an invariant relationship between the transient pharyngeal resonant, the first formant and the lower harmonics. This, according to the filters' design, is an impulse response in the amplitude domain. In contrast, glottal stops do not have an impulse response; no SPL amplitude response is indicated.

Notably, Fig's 7 and 8, similar to 9 and 10, differ from Fig's 11 and 12: in the former the presence of a phonetic glottal onset does not have amplitude response and formant resonance response. This holds in Fig's 7 and 8, though the transient pharyngeal resonant is present. In comparison, in Fig's 11 and 12, where there is no phonetic glottal present, there is an amplitude response, with and without the fundamental being present. (The filters' resonance response, FFT, was not measured here). This in-part suggests SPL amplitude blocking by the glottal: there is no natural or impulse response, acoustically. Further, the nil SPL spectrum of the filtered phonetic glottal onset, Fig's. 9 and 10, precludes mapping to complex auditory responses found to occur, particularly, in the dorsal cochlear nucleus (DCN) responses to linear phase signals in which the amplitude is odd (nonlinear). For example, in damping (Fig. 4, 650–950Hz: 3 Bark); and in sharp tuning resolution found in transient patterns involving significant SPL peaks and troughs. Moreover, the FIR filtered glottal segment's amplitude response prevents lateral inhibition.

Summarily, in contrast to the phonetic glottal onset the transient pharyngeal resonant has a unique acoustic property: relative to the frequency spectrum, the SPL amplitude response is greater in the presence of the fundamental, the lower harmonics, and the first formant - i.e., to damping in the 200-650Hz range. This can be interpreted as a transmission property, retained in the digital filtering processing. By hypothesis this property is significant in mapping to complex auditory responses, for example, sharp tuning in higher order auditory-neural functions.

Nonetheless, the findings here do not rule out the theory that sharp tuning is due to lateral inhibition (Pickles, ibid.). Also, there is no substantiation to the claim that the missing fundamental (approx. 200 Hz) alone contributes to the lack of amplitude response. As illustrated in Fig's. 7 and 9, for example, though the fundamental is present; there is no amplitude response. In contrast, Fig. 12, there is no fundamental; there is an amplitude response.

The observations rule out claims that formant resonance is filtered independently of the fundamental and lower harmonics — i.e., that in both filters high frequency energy is filtered, and thus the damping function in the 200–950Hz range, if missing, is is inconsequential as an actual resonance factor. Here, its function is invariant, albeit in binary terms, as shown by the technical filtering.

As demonstrated, when the spectrum is bandpass filtered to allow energy but no phonation or formant resonance in the low-frequency range there is nothing to which the SPL amplitude may respond or on which it may be modulated. Thus damping is prevented from occurring. This being the case is inconsequential for glottal stops; whereas resonance conditions which, present in pharyngeals, are necessary acoustically for damping. If these are absent due to filtering, digital or peripheral, one might ask how these are present - i.e., in Fig's 11 and 12 what is the amplitude response a function of.

The hypothesis here is that the amplitude response is a function of a relation between the fundamental, the lower two harmonics and the first formant. These form a resonance condition, i.e., resonance in the 200–950Hz range. The resonance condition therefore is considered here as a potential, determined by the natural impulse response, not the phonatory forced response.

4. CONCLUSION

In the preliminary design and application of the filters to the Spokane data indications in the filter responses are that the auditory system uses SPL amplitude changes of less than 3 Bark, the critical distance across formant bandwidths. While this conclusion appears profound and may be controversial, the findings are preliminary and depend on the methodological design of a systemic relationship, rather than a logical separation of transient pharyngeal transmission frequency bands from excitation period frequency bands. Using a bandpass Hamming filter it is possible to maintain this, while also obtaining the separation, resulting in a characteristic glottalic segment.

The observations here do not correspond directly to the phonemic and dependency representation of the transient pharyngeal resonant. As outlined, by tentative hypothesis the phonetic glottalic onset inhibits the transient pharyngeal - i.e., there is an acoustic-linguistic feature retained by which the two are associated, but contrastive. Further, since this relation appears ambiguous in present-day Spokane, research which might contribute to linguistic and clarification is psychoacoustic: the application of the filtered segments to obtain native speakers' judgements (p.c., B.C.Dickson). If for the native speaker the filtered segment in fact retains its contrastiveness with pharyngeals, then by hypothesis the apparent pharyngeal resonants/glottal stops alternation is phonetic, only. This could be supplemented with linguistic research into the morphology and prosodics of Spokane, with testing independent of the acoustics.

In regard to the acoustics of root-initial pharyngeal resonants in CV environments, while the amplitude generated alone by the fundamental is used in phonatory production, alone, this neither accounts for nor preserves the apparent transient damping in the low-frequency first formant region of the spectrum. Whereas FIR filter designs based on source-filter theory require a transfer function that will satisfy formant transitions expanded in terms of bandwidths and the fundamental, the same derivational basis is not present from input (microphonic) data. Expansions in the same terms therefore are not possible, in a DTLTI transmission system specific to pharyngeal resonants. At the microphonic input level the transfer function is a higher order, a form of transmission, which as a filter can be considered distinct from but not independent of the excitation (source). It should in this way be differenced, formally, and as such be expanded functionally, as an operation on damping - this in attempts to identify any correspond it has to segmentation in the auditory system. FIR filtering, with constant phase shifts, allows for this as an operation.

Further acoustic research is required to address one of the FIR's conventional disadvantages: stability of the higher-order filter number and its concomittant increasing time delays in accounting for the observed amplitude response. This could be examined, perhaps by differencing filtering and damping in the transmission transfer functions corresponding to delays in the dorsal cochlear nucleus' complex signal processing.

NOTES

The author wishes to thank D. Wong; Dr.'s B. F. Carlson and J. H. Esling; and B. C. Dickson. Digital analyses were done in the Linguistics Dept., Phonetics Laboratory, University of Victoria, Canada. Spokane tape recordings (1969, 1992), provided by B. F. Carlson, were digitized by the author using a 10k/sec. sampling rate.

REFERENCES

Carlson, B. F. (1969). Spokane, Interior Salish. Cassette Tape Recordings. Word Lists.

Carlson, B. F. (1992). Spokane, Interior Salish. Cassette Tape Recordings. Word Lists.

Dickson, B. C. and J. A. Clayards. (1990). *User's Guide to the CSL© Program.* Victoria, Canada: Speech Technology Research, and Pine Brook, N. J.: Kay Elemetrics.

Esling, J.H. and B.C. Dickson. (1991). Automatic procedure for laryngographic (Lx) analysis of phonation contrasts. *Proceedings of the XIIth International Congress of Phonetic Science* [*ICPhS*], Aix-en-France, France: Université de Provence. pp. 6–9.

Fant, G. (1970). *Acoustic Theory of Speech Production.* The Hague: Mouton.

Fant, G. (1980). The Relations between Area Functions and the Acoustic Signal. *Phonetica* 37:55 – 86.

Lagerstrom, P.A. (1988). *Matched Asymptotic Expansions.* New York: Springer-Verlag.

Laver, J. (1980). *The Phonetic Description of Voice Quality.* Cambridge: Cambridge University Press.

Maddieson, I. and Sandra Ferrari Disner (contributor). (1984). *Patterns of Sounds.* Cambridge: Cambridge University Press.

Pickles, J.O. (1991). *An Introduction to the Physiology of Hearing.* 2nd ed. London: Academic Press.

Stanley, W., G. R. Dougherty and R. Dougherty. (1984). *Digital Signal Processing.* 2nd ed. Reston, VA: Reston.

Stevens, K. N. and S. J. Keyser. (1989). Primary Features and Their Enhancement in Consonants. *Language* 65:81–106.

Stevens, K. N. and J. S. Perkell. (1977). Speech Physiology and Phonetic Features, M. Sawashima and F.S. Copper (eds.) *Dynamic Aspects of Speech Production.* Tokyo: University of Tokyo Press. pp. 323–341.

Fig. 1

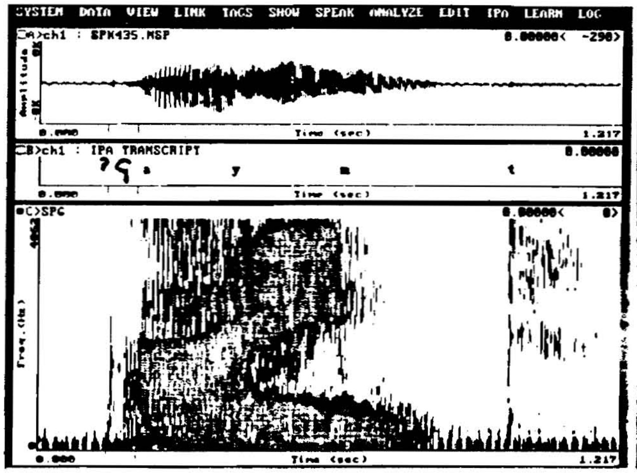

Fig. 2

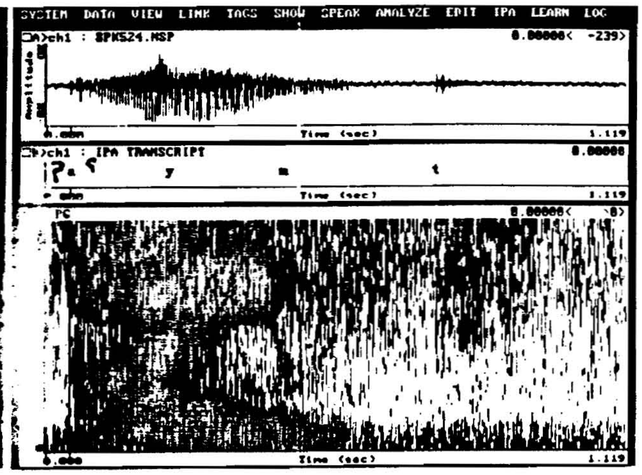

Fig. 3

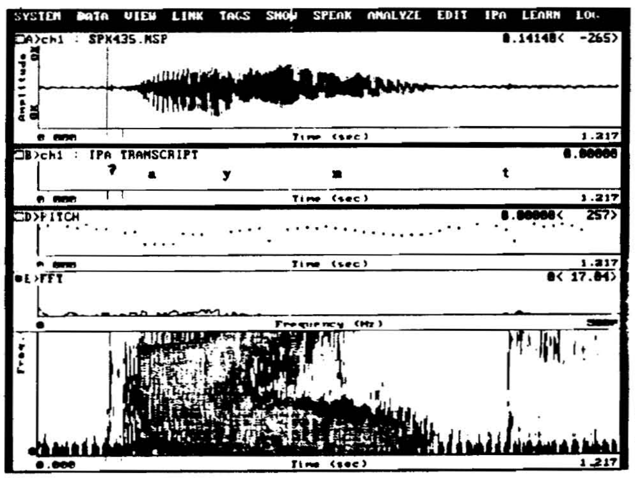

Fig. 4

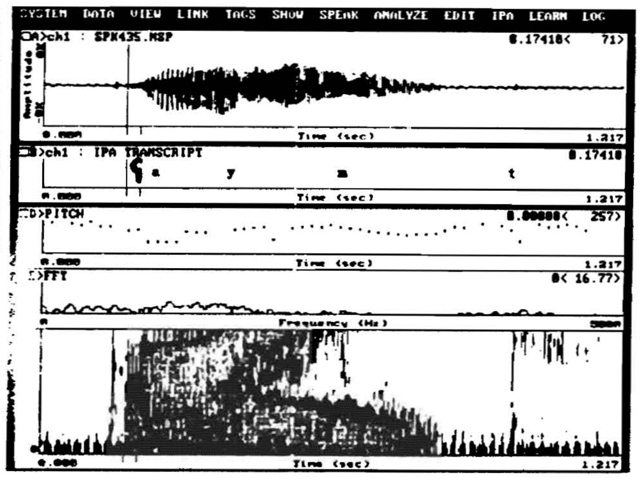

Fig. 5

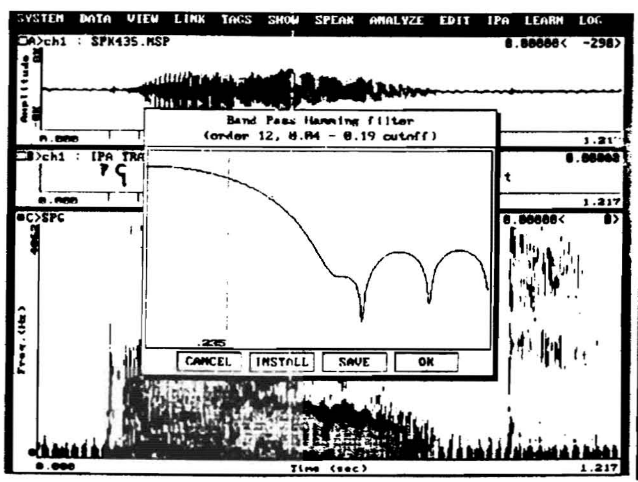

Fig. 6

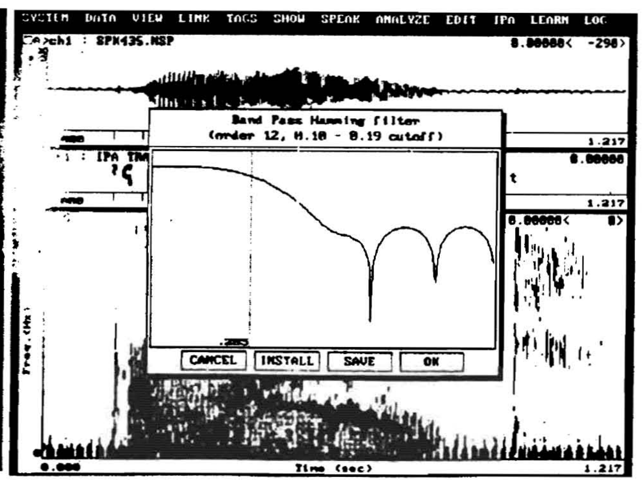

Fig. 7

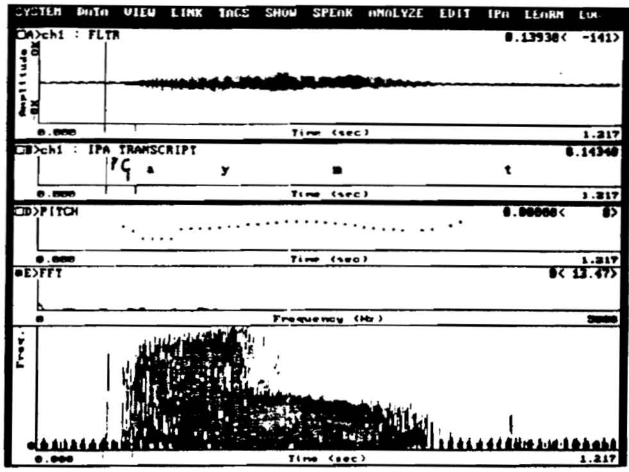

Fig. 8

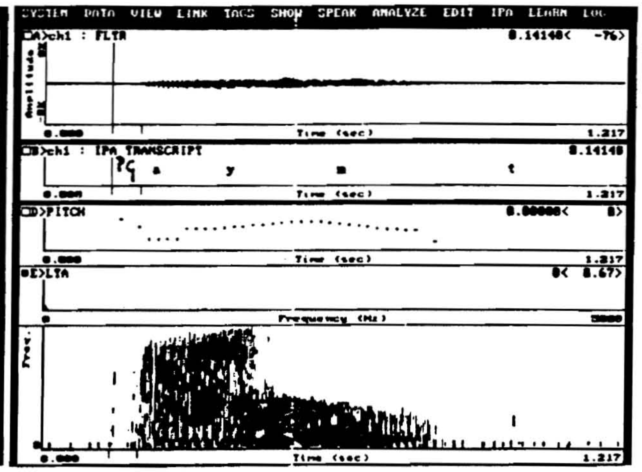

Fig. 9

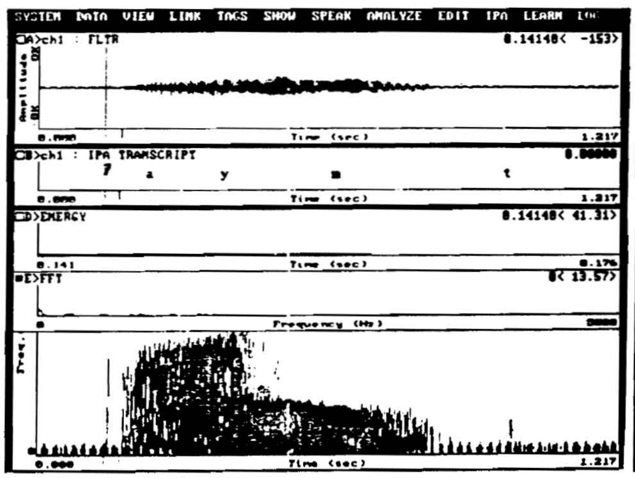

Fig. 10

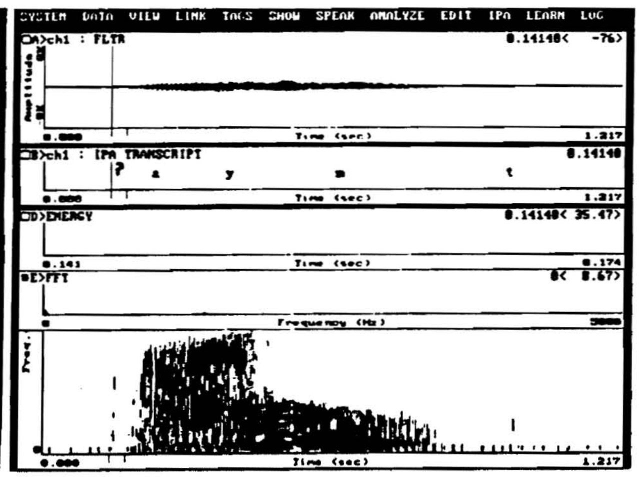

Fig. 11

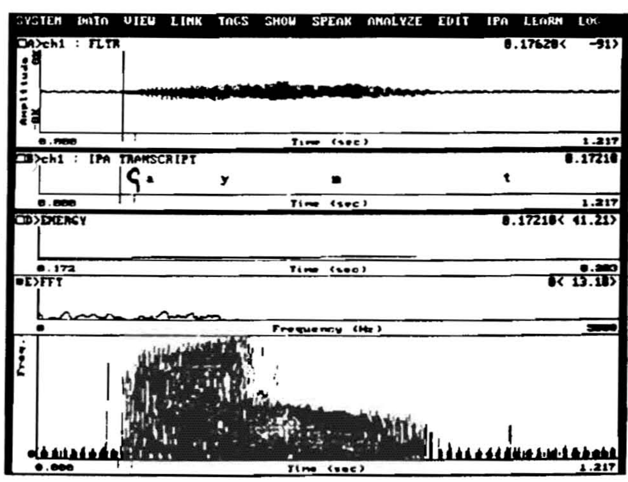

Fig. 12