APPLICATION OF *FIR* DIGITAL FILTERS TO MODEL HUMAN AUDITORY PERCEPTION OF SHORT-TIME AMPLITUDE IN LOW-FREQUENCY RAPID SPECTRUM CHANGE*

Brenda Orser
Department of Linguistics, University of Victoria
P.O.Box 3045, Victoria, B.C. V8W 3P4 Canada

Introduction: FIR (finite impulse response) filters are applied using CSL,[1,2] in the quantization of the SPL amplitude component of linguistic resonant pharyngeal consonants,/ʕ/, in word-initial CV environments, Fig.1. Only the onset consonant amplitude responses are modelled, with FIR transfer functions corresponding to SPL acoustic responses: glottal excitation (forced response), Fig.2, separate from the pharyngeal transmission (natural response), Fig.3. The FIR transfer function (impulse response) system design has the following properties: pharyngeal transmission interpreted as odd amplitude/even phase; this is differentiated from glottal source characteristics: even amplitude/odd phase. This motivates a DTLTI system.

The objective of the FIR design as applied to speech transmission is to derive linguistic structure, where transmission functions correspond to both segmentation in the auditory system, and active articulatory mechanisms in speech production. The general issue is the separation of the system from the source, decoupling vocal-cavity transfer resonances from vibrations of the vocal cords.[3] The basic design problem is the filter order: specifically, 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 for the auditory analysis of pharyngeal resonants. These, as transients, exhibit an increased and rapid spectrum change of amplitude in the low-mid frequency region, Fig.3. Also, the decreasing energy in the glottal excitation contrasts with the increasing energy exhibited in the transient[4,5] (Figures not incl.).

Filters: Neither the (temporal) rapidity of the spectral change, nor the bandwidth are modelled. These instead are derived as constants, which are linear. An FIR filter structure is implemented to obtain a linear phase and thus a constant phase shift, allowing the amplitude response to be modelled. In an FIR filter the impulse response *h(n)* is limited to a finite number of points.[6] The impulse response is expressed as

(1) $h(n) = \alpha_n$ for $0 \le n \le k$, $= 0$ elsewhere, or (2) $h(n) = \sum_{i=0}^{k} \alpha_i \delta(n - i)$

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

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

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 (or 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.[7,8] This is a systemic transmission property.

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

(5) $A(f) = w$ (6) $\beta(f) = \frac{\pi}{2} - MTw = \frac{\pi}{2} - 2\pi MTf$

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.[9] 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 only if phonatory features are added.[10,11]

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 and a low-high cutoff of .04-.13 (Figure not shown). Thus, both the fundamental and the bandwidth approximating the second formant are filtered. 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.

Summarily, the filter's input applied to the waveform gives a minimalized amplitude response, Fig. 4, D&E screens (note negative SPL value, in D). The auditory sound in E is a woody pulse - flatly-tuned and lacking resonance. This is similar perceptually to the glottal stop, [ʔ], in which there is minimal laryngeal activity, ideally -

i.e., [-voice], Fig. 2. Alternatively, broadband noise sounds, e.g., [h], can be produced by widening the bandpass cutoff and maintaining even amplitude (filters not shown). The SPL spectrum of either segment precludes mapping to complex auditory responses found in, particularly, the dorsal cochlear nucleus (DCN) responses to linear phase signals in which the amplitude is odd (nonlinear). For example, in damping (Fig.3, 650-950Hz); and in sharp tuning resolution found in transient patterns involving significant SPL peaks and troughs. Moreover, the FIR filter's amplitude response prevents lateral inhibition. Nonetheless, this does not rule out the theory that sharp tuning is due to lateral inhibition.[12] Noteably, the missing fundamental alone does not contribute to the lack of amplitude response. The damping function is missing as a factor, prevented by the technical filtering.

Conclusion: In applications of the preliminary filter design it is concluded that apparently the auditory system does use SPL amplitude changes of less than 3 Bark, the *critical distance*[13] across formant bandwidths. But this requires the separation of transmission frequency bands of and for transients, from those of excitation periods. Further, in regard to pharyngeal resonants, in word-initial CV environments, maintaining the amplitude generated by the fundamental neither accounts for nor preserves the apparent transient damping in the second formant region. 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, thus its expansions are not possible, in a DTLTI transmission system specific to pharyngeal resonants. In this, the transfer function is a higher order, a form of transmission, which as a filter can be considered separate from the excitation (source). It should therefore be expanded functionally, as an operation on damping, to thus correspond to segmentation in the auditory system. FIR filtering, with constant phase shifts, allows for this as an operation.

Further 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.

References: *The authour wishes to thank D. Wong; Dr.'s. J.H. Esling and B.F. Carlson; and B.C. Dickson. Digital analyses were done in the Linguistics Dept., Phonetics Laboratory, University of Victoria, Canada. Spokane (Interior Salish) tape recordings (1969), provided by B.F. Carlson, were digitized with sampling @ 10k/sec.
[1]Dickson, B.C.,& J.A. Clayards, 1990, *User's Guide to the CSL Program*, Speech Technology Research, Victoria, Canada. [2]Snell, R.C., 1990, *CSL Program* Signal processing software. [3]Fant, G., 1980, The Relations between Area Functions and the Acoustic Signal, *Phonetica* 37:55-86; 1970, *Acoustic Theory of Speech Production*, The Hague: Mouton. [4]Stevens, K.N.,& S.J.Keyser, 1989, Primary Features and Their Enhancement in Consonants, *Language* 65:81-106. [5]Stevens, K.N.,& J.S.Perkell, 1977, Speech Physiology and Phonetic Features, M. Sawashima & F.S. Copper (eds.) *Dynamic Aspects of Speech Prod.*, Tokyo: U. Tokyo Press: 323-341. [6]Stanley, W., Dougherty, G.R.,& R. Dougherty, 1984. *Digital Signal Processing*, 2nd ed. Reston, VA: Reston, (incl. equations (1)-(6)). [7]ibid:223. [8]Lagerstrom, P.A., 1988. *Matched Asymptotic Expansions*. NY: Springer-Verlag. [9]Stanley, et. al., ibid:218-226. [10]Laver, J. 1980. *The Phonetic Description of Voice Quality*, Camb: Cambridge UP. [11]Esling, J.H.,et.al., 1991, Automatic procedure for laryngographic (Lx) analysis of phonation contrasts, *Proc., XIIth ICPhS*, France:Univ. de Provence:6-9; [12]Pickles, J.O. 1991. *An Intro. to the Physiology of Hearing*, 2nd ed., London: A.P. [13]Chistovich, L. 1985. Central auditory processing of peripheral vowel spectra, *Jl Acous Soc Am* 38:424-428.

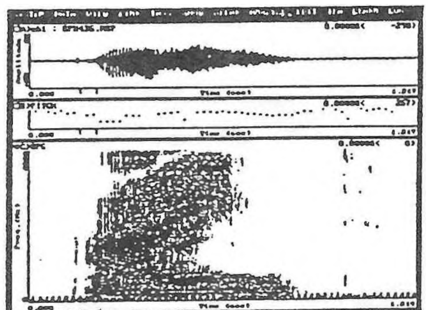

Fig. 1

/ʔaymt/

'angry'

ʔa y m t

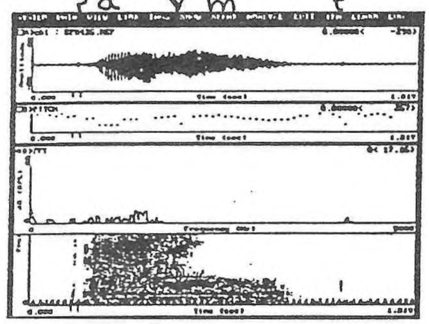

Fig. 2

[ʔ]

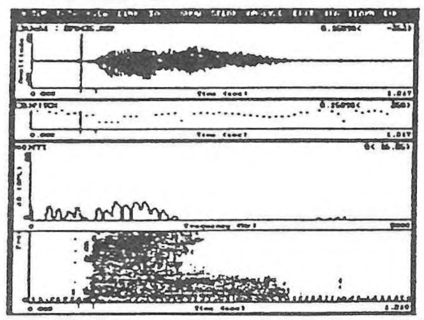

Fig. 3

[ʕ]

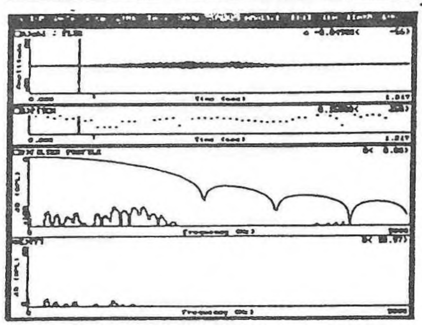

Fig. 4

D

E