An FIR Implementation of Zero Frequency Filtering of Speech Signals

Kruthiventi S. S. Srinivas and Kishore Prahallad

Abstract—**Zero frequency filtering is a technique used in the characterization and analysis of glottal activity from speech signals. The filter design originally proposed has an infinite impulse response (IIR) filter followed by two successive finite impulse response (FIR) filters. In this paper, the process of its computation is analyzed and a simplified FIR implementation is proposed by employing the inherent pole-zero cancellation involved in the process. Theoretical proofs are derived in both, frequency and time domains. We show that the theoretically derived FIR filter is a convolution of two filters, whose impulse responses are triangular shaped. The advantage of the proposed FIR filter lies in reduction of computational requirements for zero frequency filtering which include – 1) use of single-precision floating point and 2) stability of the filter.**

Index Terms—**Finite Impulse response (FIR) filter, glottal activity detection, infinite impulse response (IIR) filter, zero frequency filtering (FZF).**

I. Introduction

Zero frequency filtering (ZFF) has been shown to be effective for characterization of glottal activity [1]. It is a filter design which has the advantages of preserving and highlighting the glottal activity present in speech production [2]. It has also been applied for extraction of pitch contours from two speakers in a mixed speech signal [3]. The essential idea involved in zero frequency filtering is that an impulse like stimulus to a stable system will result in a discontinuity in the system's output. During the production of voiced speech, there occurs a periodic opening and closing of the vocal folds. The closing is so rapid that it almost results in an impulse kind of excitation to the vocal tract system. This will always result in a discontinuity in the output of the vocal tract independent of its configuration. This discontinuity is spread over the entire spectrum of its response. Zero frequency is chosen to observe the discontinuities as it is effectively shielded from time varying resonances of the vocal tract [1]. Thus, the glottal activity information is extracted using a narrow frequency band filtering of speech at zero frequency.

The filter design originally proposed for zero frequency filtering in [1] has an infinite impulse response (IIR) filter followed by two successive finite impulse response (FIR) filters. The IIR filter has four poles, and the output of the filtered signal grows as a third degree polynomial with time. The precision and the number of bits to store this polynomial output increases with the length of the signal. The two successive FIR filters in cascade (referred to as detrending filters) increase the computational requirements.

In this paper, we show that the filter design of an IIR filter followed by two detrending FIR filters as proposed in [1] can be collapsed into a single FIR filter. The proofs are derived theoretically in both time and frequency domains.

Manuscript received September 01, 2011; revised April 20, 2012; accepted June 10, 2012. Date of publication July 05, 2012; date of current version September 21, 2012. This work was done while K. S. S. Srinivas was a summer intern at IIIT Hyderabad during May-July, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Keiichi Tokuda.

K. S. S. Srinivas is with the Indian Institute of Technology Madras, Chennai 600036, India (e-mail:kssaisrinvias@gmail.com).

K. Prahallad is with the International Institute of Information Technology (IIIT), Hyderabad 500032, India (e-email:kishore@iiit.ac.in).

Digital Object Identifier 10.1109/TASL.2012.2207114

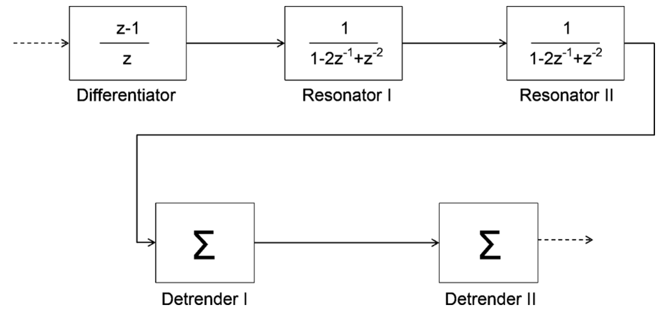

Fig. 1. Block diagram of the IIR implementation as in [1]. Here, resonator represents an IIR filter with two poles at $z = 1$ as in (2) and (3). Detrender represents removal of local mean as in (4) and (5).

The organization of this paper is as follows: Section II presents the original formulation of zero frequency filtering of speech signal. Section III derives the FIR equivalent filter in both time and frequency domains. Section IV provides a comparative analysis of IIR and FIR implementations of zero frequency filtering.

II. IIR Implementation of Zero Frequency Filtering

The steps involved in zero frequency filtering of speech signal are presented below. These steps are initially proposed in [1]. A more detailed analysis of the method is provided in [4].

1) Perform preemphasis on speech signal $s[n]$ using a difference operation.

$$x[n] = s[n] - s[n-1]. \tag{1}$$

2) Filter the speech signal through an ideal zero frequency resonator which has got four poles placed at $z = 1$ in the Z plane.

$$r_1 = x[n] + 2r_1[n-1] - r_1[n-2] \tag{2}$$
$$r[n] = r_1[n] + 2r[n-1] - r[n-2]. \tag{3}$$

3) Remove the trend in $r[n]$ by subtracting the local mean computed over a window of length $2N + 1$. This process helps in highlighting the discontinuities in the filtered signal due to impulse type of excitation. The value of window length is not critical as long as it lies in the range of one to two pitch periods.

$$f_1[n] = r[n] - \frac{1}{2N+1}\sum_{k=-N}^{N} r[n-k] \tag{4}$$

$$f[n] = f_1[n] - \frac{1}{2N+1}\sum_{k=-N}^{N} f_1[n-k]. \tag{5}$$

The block diagram of the IIR implementation is shown in Fig. 1.

III. FIR Implementation of Zero Frequency Filtering

Since each of the filters used in IIR implementation is linear time-invariant, it holds the properties of commutativity and associativity [5]. An alternative implementation which does not alter the system's output is shown in Fig. 2. We use this alternative implementation to derive the FIR filter.

A. Derivation of FIR Filter in Time Domain

To proceed further, we first consider a resonator (with two poles) and a detrender in the alternative implementation shown in Fig. 2. The difference equation representation for these two systems respectively is –

$$y[n] = x[n] + 2y[n-1] - y[n-2] \tag{6}$$

1558-7916/$31.00 © 2012 IEEE

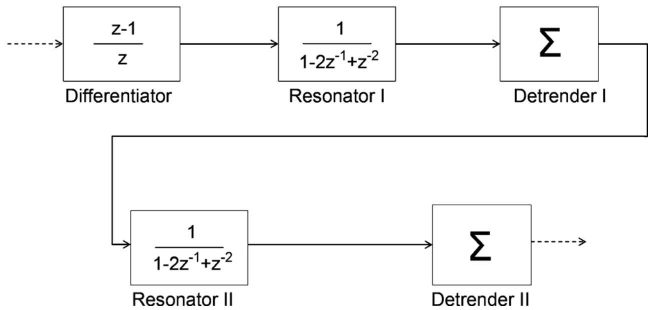

Fig. 2. An alternative implementation of the zero frequency filtering which does not alter the system's output.

$$d[n] = y[n] - \frac{1}{2N+1}\sum_{k=-N}^{N} y[n-k] \tag{7}$$

where $x[n]$ is the output of the differentiator.

$y[n]$ in (6) can be simplified as

$$y[n] - y[n-1] = x[n] + y[n-1] - y[n-2]. \tag{8}$$

$$\text{Let } y_1[n] = y[n] - y[n-1]. \tag{9}$$

Substitute (9) in (8),

$$\begin{aligned} y_1[n] &= x[n] + y_1[n-1] \\ &= \sum_{m=0}^{\infty} x[n-m]. \end{aligned} \tag{10}$$

From (9)

$$\begin{aligned} y[n] &= y_1[n] + y[n-1] \\ &= \sum_{l=0}^{\infty} y_1[n-l] \\ &= \sum_{l=0}^{\infty}\sum_{m=0}^{\infty} x[n-l-m] \\ &= \sum_{k=0}^{\infty}(k+1)x[n-k] \end{aligned}$$

which implies $y[n]$ can be expanded as the following infinite series.

$$y[n] = x[n] + 2x[n-1] + 3x[n-2] + \cdots \infty. \tag{11}$$

Substituting (11) into (7) term-wise and expanding

$$\begin{aligned} &d[n] = \\ &x[n] - \frac{1}{2N+1}\left(x[n+N] + \cdots + x[n-N]\right) \\ &+ 2x[n-1] - \frac{2}{2N+1}(x[n+N-1] + \cdots + x[n-N-1]) \\ &+ 3x[n-2] - \frac{3}{2N+1}(x[n+N-2] + \cdots + x[n-N-2]) \\ &. \\ &. \\ &+ (m+1)x[n-m] \\ &- \frac{m+1}{2N+1}\left(x[n+N-m] + \cdots + x[n-N-m]\right) \\ &. \\ &. \\ &\infty. \end{aligned} \tag{12}$$

The summation can be simplified by collecting the individual coefficients of each of the delay terms from all summations and adding them up.

$$\begin{aligned} &d[n] \\ &= -\sum_{m=1}^{N}\left(\frac{\sum_{k=1}^{N-m+1} k}{2N+1}\right)x[n+m] \\ &+ \sum_{m=0}^{N-1}\left(m+1-\frac{\sum_{k=1}^{m+N+1} k}{2N+1}\right)x[n-m] \\ &+ \sum_{m=N}^{\infty}\left(m+1-\overbrace{\frac{\sum_{k=m-N+1}^{m+N+1} k}{2N+1}}^{\Gamma}\right)x[n-m]. \end{aligned} \tag{13}$$

In the above series, the coefficients of terms with $m \geq N$ can be simplified as shown below:

$$\begin{aligned} \text{Let } \Gamma &= \frac{1}{2N+1}\sum_{k=m-N+1}^{m+N+1} k \\ &= \frac{1}{2N+1}\left(\sum_{k=1}^{m+N+1} k - \sum_{k=1}^{m-N} k\right) \\ &= \frac{1}{2N+1}\left(\tfrac{(m+N+1)(m+N+2)}{2} - \tfrac{(m-N)(m-N+1)}{2}\right) \\ &= (m+1). \end{aligned} \tag{14}$$

On substitution of (14) back in (13), $d[n]$ is reduced to a finite series. Further, using the sum of m natural numbers, $d[n]$ can be written as follows.

$$\begin{aligned} d[n] =& -\sum_{m=1}^{N}\left(\frac{(N-m+1)(N-m+2)}{2(2N+1)}\right)x[n+m] \\ &+\sum_{m=0}^{N-1}\left(m+1-\frac{(m+N+1)(m+N+2)}{2(2N+1)}\right)x[n-m] \\ =& \frac{-1}{2N+1}\sum_{m=1}^{N}\left(\frac{(N-m+1)(N-m+2)}{2}\right)x[n+m] \\ &- \frac{1}{2N+1}\sum_{m=0}^{N-1}\left(\frac{(N-m-1)(N-m)}{2}\right)x[n-m] \\ =& \frac{-1}{2N+1}\sum_{m=1}^{N}\left(T_{N-m+1}\right)x[n+m] \\ &- \frac{1}{2N+1}\sum_{m=0}^{N-1}\left(T_{N-m-1}\right)x[n-m] \end{aligned} \tag{15}$$

where T_m denotes the m^{th} triangular number which is the sum of first m natural numbers, i.e.,

$$T_m = (1 + 2 + \cdots + m) = \frac{m(m+1)}{2}. \tag{16}$$

The two summations in (15) can be combined using a new index r to include both the positive and negative delay terms of $x[n]$.

Rewriting (15),

$$d[n] = \frac{-1}{2N+1}\left(\sum_{r=0}^{2N-2} x[(n+N)-r]t[r]\right) \tag{17}$$

where $t[r]$ is given as

$$t[r] = \begin{cases} T_{r+1} & \text{if } r < N, \\ T_{(2N-1)-r} & \text{if } r \geq N, \end{cases}. \tag{18}$$

From the definition of convolution,

$$\begin{aligned} d[n] &= \frac{-1}{2N+1} t[n] * x[n+N] \\ &= \frac{-1}{2N+1}\{T_1 \dots T_{N-1}\ T_N\ T_{N-1} \dots T_1\} \\ &\quad * x[n+N] \end{aligned} \tag{19}$$

where $*$ denotes the convolution operator and $\{T_1 \dots T_{N-1}\ T_N\ T_{N-1} \dots T_1\}$ denotes the time-series representation of the signal $t[n]$.

In Fig. 2, a resonator and a detrender are cascaded twice, the total system response (apart from the differentiator) is obtained by

$$\frac{\{(T_1 \dots T_N \dots T_1) * (T_1 \dots T_N \dots T_1) * x[n+N]\}}{(2N+1)^2}. \tag{20}$$

B. Derivation of FIR Filter in Frequency Domain

Let the Z transforms of $x[n], y[n]$ and $d[n]$ respectively be $X(z), Y(z)$ and $D(z)$. The Z transforms of the two pole resonator in (6) and the detrending filter in (7) respectively are

$$\frac{Y(z)}{X(z)} = \frac{1}{1 - 2z^{-1} + z^{-2}} = \frac{1}{(1 - z^{-1})^2} \tag{21}$$

$$\frac{D(z)}{Y(z)} = \frac{-1}{2N+1}\left(z^{-N} + \dots + z^{-1} - 2N + z + \dots + z^{N}\right). \tag{22}$$

The summation in the parentheses can be factorized as

$$z^{-N} + \dots + z^{-1} - 2N + z + \dots + z^{N} = \left(z - 2 + z^{-1}\right) f_\Delta(N, z). \tag{23}$$

A proof by induction for (23) is given in Appendix. Here $f_\Delta(N, z)$ is a polynomial in z domain with triangular numbers as its symmetric coefficients.

$$\begin{aligned} f_\Delta(N, z) = T_1 z^{N-1} + \dots + T_{N-1} z \\ + T_N + T_{N-1} z^{-1} + \dots + T_1 z^{-N+1}, \end{aligned} \tag{24}$$

where T_m is given by (16).

From (22) and (23),

$$\begin{aligned} \frac{D(z)}{Y(z)} &= \frac{-\left(z - 2 + z^{-1}\right)}{2N+1} f_\Delta(N, z) \\ &= \frac{-z(1 - z^{-1})^2}{2N+1} f_\Delta(N, z). \end{aligned} \tag{25}$$

The Z transform of the resonator in cascade with trend removal filter is given by the product of (21) and (25).

$$\begin{aligned} \frac{D(z)}{Y(z)}\frac{Y(z)}{X(z)} &= \frac{1}{(1 - z^{-1})^2}\frac{-z(1 - z^{-1})^2}{2N+1} f_\Delta(N, z) \\ \frac{D(z)}{X(z)} &= \frac{-z}{2N+1} f_\Delta(N, z). \end{aligned} \tag{26}$$

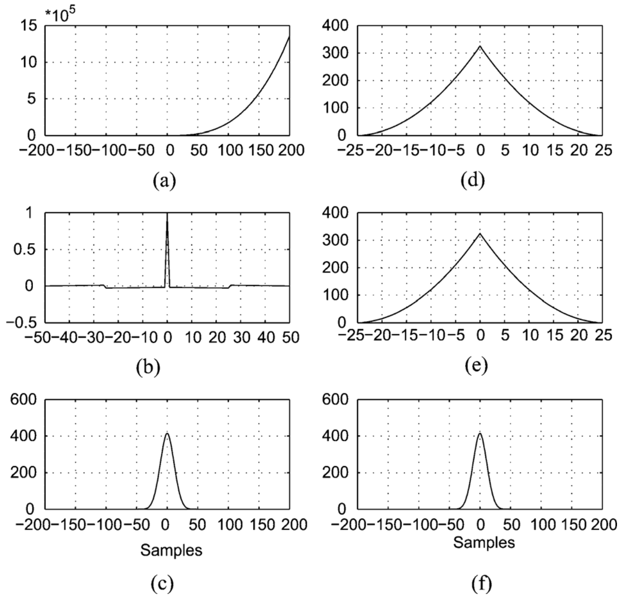

Fig. 3. Time domain response of (a) four pole resonator (b) double trend removal filter (c) zero frequency filter with IIR implementation (d) cascade of two-pole pole resonator and a detrender (e) cascade of two-pole pole resonator and a detrender (f) zero frequency filter with FIR implementation.

Since there are two such systems in cascade in Fig. 2, their combined system response is obtained by the square of (26).

$$\begin{aligned} \frac{F(z)}{X(z)} &= \left(\frac{D(z)}{X(z)}\right)^2 \\ &= \frac{z^2}{(2N+1)^2}(f_\Delta(N, z))^2 \\ &= \frac{z^2}{(2N+1)^2}(T_1 z^{N-1} + \dots + T_N + \dots + T_1 z^{-N+1})^2. \end{aligned} \tag{27}$$

Equivalent representation of the above filter is obtained by its inverse Z transform

$$\frac{1}{(2N+1)^2}\{(T_1 \dots T_N \dots T_1) * (T_1 \dots T_N \dots T_1)\}. \tag{28}$$

This is in agreement with the (20), and it is the implementation of zero frequency filtering as an FIR filter.

IV. Comparison of FIR and IIR Implementations

The zero frequency resonator with four poles is expected to grow as a third degree polynomial, in magnitude with time. The response of the four pole resonator in the time domain is shown in Fig. 3(a).

The double trend removal filter is a special kind of differentiator which effectively removes the growing trend present in the signal Fig. 3(a) and highlights the subtle fluctuations caused due to impulse excitation to the resonator. The impulse response of this double trend removal filter is shown in Fig. 3(b) for a length of $2N + 1 = 51$. The combined response of the four pole resonator in cascade with the two trend removal filters will result in a rise and fall at the instant of impulse excitation to the resonator. This acts like a smoothing filter with the weighted mean computed over approximately twice the length of the detrender. The system response of zero frequency filtering for a detrending length of $2N + 1 = 51$, with an IIR filter in time domain is shown in Fig. 3(c).

TABLE I
COMPARISON OF TIME AND SPACE REQUIREMENTS FOR FIR AND IIR IMPLEMENTATIONS

	FIR Implementation	**IIR Implementation**
Stable	yes	no
Multiplications per input sample	117	124
Additions per input sample	116	124
Precision required	single-precision	double-precision or higher

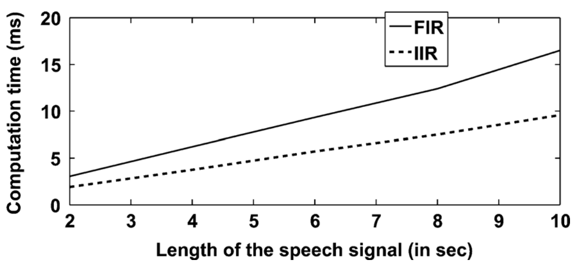

Fig. 4. Performance rest for zero frequency filtering using IIR and FIR filters.

The time domain response of the proposed FIR filter as given in (19), which is derived from the cascade of a two pole resonator and a detrender is shown in Fig. 3(d). Following (20), this filter is convolved with itself (Fig. 3(e)). The resulting output is smoother and its duration is doubled. The time domain response of the FIR filter in (20) is shown in Fig. 3(f). This is identical to the impulse response of the IIR filter cascaded with double trend removal filter as shown in Fig. 3(c). Thus the glottal closure instants from the proposed FIR implementation (Section III) match exactly with the IIR implementation (Section II).

A. Time and Space Complexity

A realization of the IIR implementation of zero frequency filtering with a detrender of length $2N+1$ requires $4N+4$ additions and $4N+4$ multiplications whereas the FIR implementation requires $4N-4$ additions and $4N-3$ multiplications. Even though there is not much reduction in the number of additions and multiplications, the computational time required differs significantly due to the high precision computation needed for the IIR implementation. A comparison of the IIR and FIR implementations is presented below for detrender length $2N+1=61$ in Table I.

As shown in the Fig. 4, the time taken for computation of zero frequency filtering through IIR is approximately 1.65 times that of the FIR.

In IIR implementation, output of the four-pole resonator grows as a third degree polynomial with time. The essential speech information is present in the form of subtle fluctuations over this ever increasing output. The two successive detrending filters need to perform high precision filtering to capture these subtle fluctuations and highlight them in the form of a zero frequency filtered signal. Due to this demand on computation, the IIR form cannot be implemented using single-precision (32-bit) arithmetic unlike the FIR implementation. This has been verified by using single-precision variables in MATLAB and it has been found that all the essential speech information which is in the form of subtle fluctuations is lost.

Further the IIR implementation involves a filter with four poles on the unit circle; the system doesn't have BIBO (Bounded Input-Bounded Output) stability which can result in spikes of large amplitudes in the filtered signal.

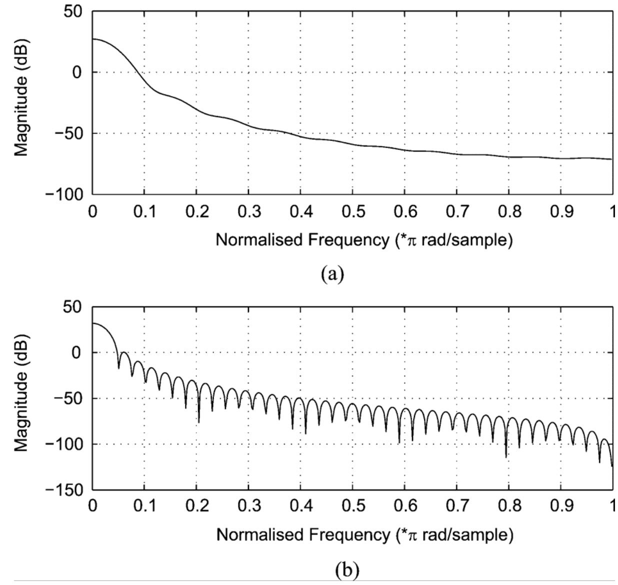

Fig. 5. Magnitude response of (a) Zero frequency filtering with FIR of length 77. (b) Hann window of length 77.

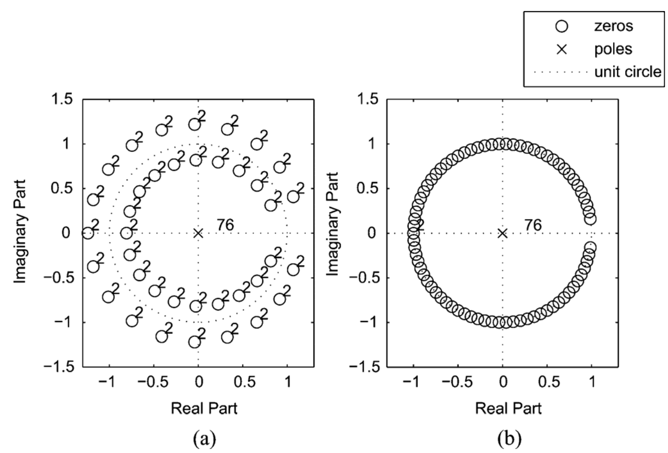

Fig. 6. Pole-Zero plots of (a) Zero frequency filtering with FIR of length 77. (b) Hann window of length 77.

V. DISCUSSION AND CONCLUSIONS

In this paper, we have shown an FIR implementation of the zero frequency filtering technique proposed originally with an IIR filter cascaded with two trend removal filters. The FIR equivalence is derived theoretically in both time and frequency domains.

The key benefits of an FIR over an IIR implementation for zero-frequency filtering (ZFF) of speech signals are – 1) computation using single-precision floating point and 2) the stability of the filter. The FIR implementation of zero-frequency filtering of a speech signal of an arbitrary length requires only single-precision floating point. For an IIR implementation of ZFF, double precision or higher precision floating point is required for arbitrary long signals (for example, signals sampled at 16 K Hz and longer than 30 seconds). Moreover, the output of an IIR filter in not always bounded.

The frequency response of the proposed FIR filter is also worth mentioning here. The FIR filter in (20) has a frequency response as shown in Fig. 5(a), with no side lobes unlike the standard Hamming or Hann windows, even though it is of a finite duration. So, using it as a window for FIR filter design will result in a system with no ripples in pass/stop bands at the expense of a larger transition width. This property can be explained from its pole-zero plot presented in Fig. 6(b). The system is a mixed-phase FIR system with zeros symmetrically placed around the unit circle both inside and outside of it, with one as the inverse of the other in magnitude. Had there been zeros on the unit circle it would have resulted in a frequency response with side lobes similar to the Hamming or Hann windows. One could also explore the FIR filter in (20) to study short-time spectral properties of speech signals including formant extraction.

Applications of zero-frequency filtering of speech signals include voiced/unvoiced detection in noisy environment, speaker tracking in multi-speaker environments [6], [7] . Such applications could be easily enabled/developed on low-end computing devices such as smart phones or hand-held devices using the FIR implementation of ZFF.

APPENDIX
PROOF FOR (23) BY THE PRINCIPLE OF MATHEMATICAL INDUCTION

Let

$$P(n) : z^{-n} + \cdots + z^{-1} - 2n + z + \cdots + z^{n} = \left(z - 2 + z^{-1}\right) f_{\Delta}(n, z) \quad (29)$$

Basis: $P(1)$ amounts to the statement

$$z^{-1} - 2 + z^{1} = \left(z - 2 + z^{-1}\right) f_{\Delta}(1, z). \quad (30)$$

Consider the right-hand side of the equation.

$$\begin{aligned}\left(z - 2 + z^{-1}\right) f_{\Delta}(1, z) &= \left(z - 2 + z^{-1}\right) T_1 \\ &= z - 2 + z^{-1}\end{aligned}$$

which is the same as the left-hand side of (30). Hence $P(n)$ is proved to be true for $n = 1$.

Inductive Step:

Assume $P(k)$ holds true for some natural number k. So

$$z^{-k} + \cdots + z^{-1} - 2k + z + \cdots + z^{k} = \left(z - 2 + z^{-1}\right) f_{\Delta}(k, z). \quad (31)$$

It has to be shown that $P(k+1)$ holds true, that is:

$$z^{-(k+1)} + \cdots + z^{-1} - 2(k+1) + z + \cdots + z^{k+1} = \left(z - 2 + z^{-1}\right) f_{\Delta}(k+1, z). \quad (32)$$

Consider the right-hand side of the (32)

$$\begin{aligned}&\left(z - 2 + z^{-1}\right) f_{\Delta}(k+1, z) \\ &\quad = \left(z - 2 + z^{-1}\right) \left\{ \begin{array}{c} T_1 z^{k} + T_2 z^{k-1} + \cdots + T_k z + T_{k+1} \\ + T_k z^{-1} + \cdots + T_2 z^{-k+1} + T_1 z^{-k} \end{array} \right\}.\end{aligned}$$

By its definition (16), $T_{m+1} = T_m + (m+1)$ for any natural number m. We can rewrite the above equation as

$$\begin{aligned}&\left(z - 2 + z^{-1}\right) \left\{ \begin{array}{c} T_1 z^{k-1} + T_2 z^{k-2} + \cdots + T_{k-1} z + T_k \\ + T_{k-1} z^{-1} + \cdots + T_2 z^{-k+2} + T_1 z^{-k+1} \end{array} \right\} \\ &\quad + \left(z - 2 + z^{-1}\right) \left\{ \begin{array}{c} z^{k} + 2z^{k-1} + \cdots + kz + (k+1) \\ + kz^{-1} + \cdots + 2z^{-k+1} + 1z^{-k} \end{array} \right\}.\end{aligned}$$

From (31)

$$\begin{aligned}&= z^{-k} + \cdots + z^{-1} - 2k + z + \cdots + z^{k} \\ &\quad + \left(z - 2 + z^{-1}\right) \left\{ \begin{array}{c} z^{k} + 2z^{k-1} + \cdots + kz + (k+1) \\ + kz^{-1} + \cdots + 2z^{-k+1} + z^{-k} \end{array} \right\}\end{aligned}$$

Multiplying and cancelling the like terms in second term of the addition

$$\begin{aligned}&= z^{-k} + \cdots + z^{-1} - 2k + z + \cdots + z^{k} + \left(z^{-k-1} - 2 + z^{k+1}\right) \\ &= z^{-(k+1)} + z^{-k} + \cdots + z^{-1} - 2k + z + \cdots + z^{k} + z^{k+1}\end{aligned}$$

which is same as the right-hand side of (32).

Hence by the principle of induction, $P(n)$ is true for any natural number n.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for useful feedback with suggestions, which helped in refining the draft significantly.

REFERENCES

[1] K. S. R. Murty and B. Yegnanarayana, “Epoch extraction from speech signals,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1602–1613, Nov. 2008.

[2] K. S. R. Murty, B. Yegnanarayana, and A. Joseph, “Characterization of glottal activity from speech signals,” *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 469–472, Jun. 2009.

[3] B. Yegnanarayana and S. R. M. Prasanna, “Analysis of instantaneous F0 contours from two speakers mixed signal using zero frequency filtering,” in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP-2010)*, Dallas, TX, Mar. 2010, pp. 5074–5077.

[4] K. S. R. Murthy, “Significance of excitation source information for speech analysis,” Ph.D., Dept. of Comput. Sci. and Eng., IIT, Madras, Chennai, India, 2009.

[5] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, 1st ed. N. Delhi, India: Pearson Education, 1978.

[6] N. Dhananjaya and B. Yegnanarayana, “Voiced/nonvoiced detection based on robustness of voiced epochs,” *IEEE Signal Process. Lett.*, vol. 17, no. 3, pp. 273–276, Mar. 2010.

[7] B. Yegnanarayana, R. K. Swamy, and K. S. R. Murty, “Determining mixing parameters from multispeaker data using speech-specific information,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1196–1207, Aug. 2009.