Disclaimer

- The material provided in this document is not my original work and is a summary of some one else's work(s).
- A simple Google search of the title of the document will direct you to the original source of the material.
- I do not guarantee the accuracy, completeness, timeliness, validity, non-omission, merchantability or fitness of the contents of this document for any particular purpose.
- Downloaded from najeebkhan.github.io

MLSA Digital Filter

Outline

- Introduction
- Approximate Representation of the Mel Scale
  - S domain to Z domain Conversion
  - Z domain to Z domain Conversion
  - Frequency Warping
- Filter with Exponential Transfer function
- Pade Approximation
- Modification of Basic Transfer function
- Realization

Introduction

- Lower order Mel cepstrum coefficients can be used to synthesize speech with satisfactory quality
- The Mel log spectrum is approximated by a recursive filter
- Strube in [1] proposed a recursive filter for nonlinear linear predication
- Strube's filter has high sensitivity and wider distributions
- In this paper another filter is formulated which has better sensitivity and narrow distributions

Approximate Representation of the Mel Scale[2]

- S to Z domain mapping:

$$g(t) = \int_{-\infty}^{+\infty} f(\tau)h(t-\tau)d\tau \triangleq f(t) \otimes h(t) \qquad g_n = \sum_{k=-\infty}^{+\infty} f_k h_{n-k} \triangleq f_n * h_n$$

$$x_n + cy_n \Rightarrow x(t) + cy(t) \qquad f_n = \sum_{k=-\infty}^{+\infty} f_k \delta_{n-k} \Rightarrow f(t) = \sum_{k=-\infty}^{+\infty} f_k \phi_k(t)$$

$$F_L(s) = \sum_{n=-\infty}^{+\infty} f_n \Phi_n(s).$$

Approximate Representation of the Mel Scale[2]

- S to Z domain mapping:

$$G_L(s) = F_L(s)H_L(s).$$

Represents a Continous Convolution

$$\sum_{n=-\infty}^{+\infty} g_n\Phi_n(s) = \sum_{r=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} f_k h_r \Phi_r(s)\Phi_k(s).$$

$$n = r + k$$

$$\sum_{n=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} f_k h_{n-k}\Phi_n(s) = \sum_{n=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} f_n h_{n-k}\Phi_{n-k}(s)\Phi_k(s).$$

$$\Phi_n(s) = \Phi_{n-k}(s)\Phi_k(s) \qquad \Phi_n(s)\Phi_m(s) = \Phi_{n+m}(s)$$

$$\Phi_{n+1}(s) - \Phi_1(s)\Phi_n(s) = 0$$

$$\Phi_n(s) = c[\Phi_1(s)]^n.$$

$$\Phi_n(s) = [\Phi_1(s)]^n.$$

Approximate Representation of the Mel Scale[2]

- S to Z domain mapping:
- If the previous condition is satisfied then $f(t) = \sum_{k=-\infty}^{+\infty} f_k \phi_k(t)$

can be thought of as a mapping from s to z

$$F_D(z) = \sum_{n=-\infty}^{+\infty} f_n z^{-n}.$$

$$F_L(s) = \sum_{n=-\infty}^{+\infty} f_n \left\{ \left[\Phi_1(s) \right]^{-1} \right\}^{-n}$$

$$z = \left[\Phi_1(s) \right]^{-1} \triangleq M(s)$$

Approximate Representation of the Mel Scale[2]

- Z to Z domain mapping:

$$f(t) = \sum_{n=-\infty}^{\infty} f_n \phi_n(t) \qquad f(t) = \sum_{k=-\infty}^{\infty} g_k \lambda_k(t)$$

$$\lambda_k(t) = \sum_{n=-\infty}^{\infty} \psi_{k,n} \phi_n(t)$$

$$\sum_{n=-\infty}^{\infty} f_n \phi_n(t) = \sum_{k=-\infty}^{\infty} g_k \lambda_k(t) = \sum_{k=-\infty}^{\infty} g_k \sum_{n=-\infty}^{\infty} \psi_{k,n} \phi_n(t)$$

$$\sum_{n=-\infty}^{\infty} f_n \phi_n(t) = \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} g_k \psi_{k,n} \phi_n(t)$$

$$f_n = \sum_{k=-\infty}^{\infty} g_k \psi_{k,n}$$

Approximate Representation of the Mel Scale[2]

- Z to Z domain mapping:
- The properties required of the set of sequences $\{\psi_k(z)\}$ in order that convolution is preserved can be determined in a manner analogous to s to z mapping

$$\psi_r(z)\psi_{k-r}(z) = \psi_k(z)$$

$$\psi_k(z) = \left[\psi_1(z)\right]^k$$

Approximate Representation of the Mel Scale[2]

- Z to Z domain mapping:

$$f_n = \sum_{n=-\infty}^{\infty} g_k \psi_{k,n}$$

$$F(z) = \sum_{k=-\infty}^{\infty} g_k \psi_k(z)$$

$$\sum_{k=-\infty}^{\infty} f_k z^{-k} = \sum_{k=-\infty}^{\infty} g_k \left[\{\psi_1(z)\}^{-1} \right]^{-k}$$

$$Z = \{\psi_1(z)\}^{-1} = m(z)$$

$$F(z) = G(m(z))$$

Thus changing from the representation f_n to g_k is equivalent to mapping the complex variable plane z for f_n to the Z' for g_k by means of a substitution of variables z'=m(z)

Frequency Warping[2]

- Letting Ω be angular frequency in the z plane and Ω’ be angular frequency in the z’ plane
- We want z’=m(z) to satisfy

$$e^{j\Omega} = m\left[e^{j\Omega}\right] \qquad \Omega = \theta\left(\Omega\right)$$

- To satisfy this and the convolution preservation condition we have

$$\psi_k\left(e^{j\Omega}\right) = e^{-jk\theta(\Omega)}$$

Frequency Warping[2]

- Thus the Fourier transform of $\psi_k(n)$ must have an all-pass characteristic
- The negative of the phase will then correspond to the mapping between the frequency axes

$$\Psi_k(z) = \left(\frac{z^{-1} - a^*}{1 - az^{-1}}\right)^k$$

$$\hat{\Omega} = \theta(\Omega) = \arctan\left[\frac{(1 - a^2)\sin\Omega}{(1 + a^2)\cos\Omega - 2a}\right]$$

Mel Cepstrum

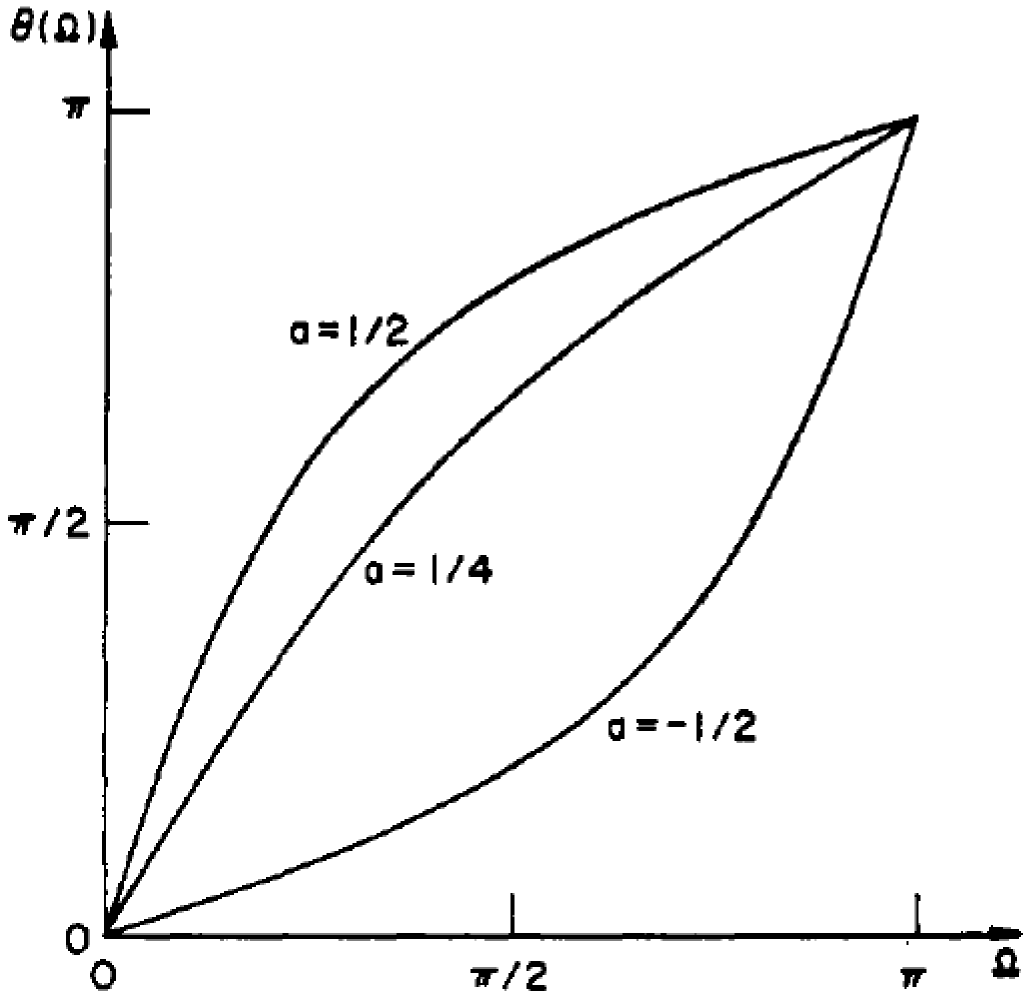
$$G_\alpha(\tilde{\Omega}) = G_0(\beta_\alpha^{-1}(\tilde{\Omega}))$$

$$\beta_\alpha^{-1}(\tilde{\Omega}) = \tan^{-1}\frac{(1-\alpha^2)\sin\tilde{\Omega}}{(1+\alpha^2)\cos\tilde{\Omega}+2\alpha}$$

$$g_\alpha(m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} G_\alpha(\tilde{\Omega})\, e^{jm\tilde{\Omega}}\, d\tilde{\Omega} = g_\alpha(-m)$$

$$c_\alpha(m) = \begin{cases} 2g_\alpha(m) & (m > 0) \\ g_\alpha(m) & (m = 0) \\ 0 & (m < 0) \end{cases}$$

Mel Cepstrum



Filter with Exponential Transfer function[3]

- To approximate the log magnitude response on the mel scale we use a filter with exponential transfer function

$$H_\alpha(\tilde{z}) = \exp\left(F_\alpha(\tilde{z})\right)$$

$$F_\alpha(\tilde{z}) = \sum_{m=0}^{M} c_\alpha(m)\,\tilde{z}^{-m}$$

$$\ln|H_\alpha(e^{j\tilde{\Omega}})| = \sum_{m=0}^{M} c_\alpha(m)\cos(m\tilde{\Omega})$$

Pade Approximation

- The general pade approximation doesn’t matches with the formulas given in the papers
- However coding the formulas in software approximates the exponential function
- We will discuss it at the end

Modification of Basic Transfer function[1]

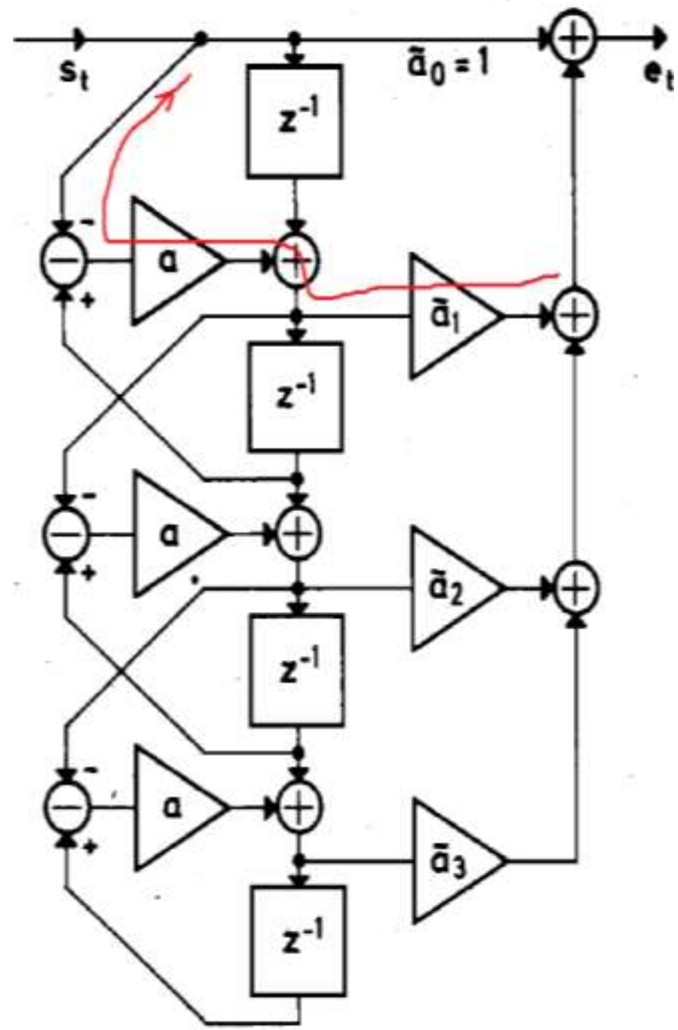
$$\bar{A}(z) = \sum_{k=0}^{p} \tilde{a}_k \bar{z}^{-k}(z), \quad \tilde{a}_0 = 1\,;$$

$$\bar{z}^{-1} = (1 - a^2) z^{-1} / (1 - a z^{-1}) - a$$

The all-pass filter is transformed to a lowpass filter

$$\bar{A}(z) = \sum_{k=0}^{p} b_k z^{-k} (1 - a z^{-1})^{-k}.$$

$$b_k = \sum_{n=k}^{p} C_{kn} \tilde{a}_n, \quad C_{kn} = \binom{n}{k} (1 - a^2)^k \, (-a)^{n-k}.$$



Modification of Basic Transfer function[1]

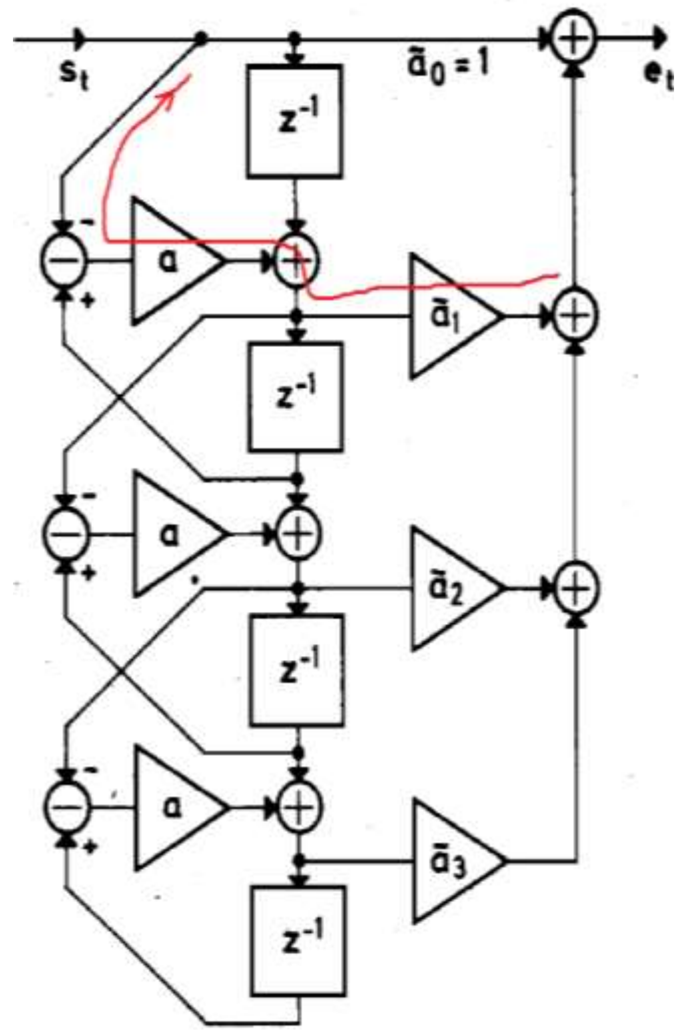
$$\bar{A}(z)=\sum_{k=0}^{p}\tilde{a}_k\tilde{z}^{-k}(z),\quad \tilde{a}_0=1\,;$$

$$\tilde{z}^{-1}=(1-a^2)z^{-1}/(1-az^{-1})-a$$

The all-pass filter is transformed to a lowpass filter

$$\bar{A}(z)=\sum_{k=0}^{p}b_k z^{-k}(1-az^{-1})^{-k}.$$

$$b_k=\sum_{n=k}^{p}C_{kn}\tilde{a}_n,\quad C_{kn}=\binom{n}{k}(1-a^2)^k\,(-a)^{n-k}.$$



$$A(z) = \sum_{k=0}^{p} a_k \tilde{z}^{-k}(z).$$

$$\tilde{z}^{-1} = \frac{(1-a^2)z^{-1}}{(1-az^{-1})} - a.$$

$$A(z) = \sum_{k=0}^{p} a_k \left[\frac{(1-a^2)z^{-1}}{(1-az^{-1})} - a \right]^k$$

The

$$A(z) = \sum_{k=0}^{p} z^{-k}(1-az^{-1})^{-k} \left[(1-a^2) - \frac{a(1-az^{-1})}{z^{-1}}\right]^k a_k.$$

$$A(z) = \sum_{k=0}^{p} b_k z^{-k}(1-az^{-1})^{-k}.$$

$$b_k = \sum_{n=}^{p}$$

$$b_k = \left[(1-a^2) - \frac{a(1-az^{-1})}{z^{-1}}\right]^k a_k.$$

e_t

Friday, Saturday, Sunday

$$A(z) = \sum_{k=0}^{p} a_k \tilde{z}^{-k}(z).$$

$$\tilde{z}^{-1} = \frac{(1-a^2)z^{-1}}{(1-az^{-1})} - a.$$

$$A(z) = \sum_{k=0}^{p} a_k \left[\frac{(1-a^2)z^{-1}}{(1-az^{-1})} - a\right]^k$$

The

$$A(z) = \sum_{k=0}^{p} z^{-k}(1-az^{-1})^{-k}\left[(1-a^2) - \frac{a(1-az^{-1})}{z^{-1}}\right]^k a_k.$$

$$b_k = \sum$$

$$A(z) = \sum_{k=0}^{p} b_k z^{-k}(1-az^{-1})^{-k}.$$

$$b_k = \left[(1-a^2) - \frac{a(1-az^{-1})}{z^{-1}}\right]^k a_k.$$

e_t

$$
\begin{aligned}
\tilde{A}(\tilde{z}) &= \sum_{n=0}^{p} \tilde{a}_n \tilde{z}^{-n}(z) \\
&= \sum_{n=0}^{p} \tilde{a}_n \Big((1-a^2)\, z^{-1} / (1-az^{-1}) - a \Big)^n \\
&= \sum_{n=0}^{p} \tilde{a}_n \sum_{k=0}^{n} \binom{n}{k} (1-a^2)^k \, z^{-k} (1-az^{-1})^{-k} \, (-a)^{n-k} \\
&= \sum_{k=0}^{p} \left[\sum_{n=k}^{p} \tilde{a}_n \binom{n}{k} (1-a^2)^k (-a)^{n-k} \right] z^{-k} (1-az^{-1})^{-k} \\
&= \sum_{k=0}^{p} b_k \, z^{-k} (1-az^{-1})^{-k}
\end{aligned}
$$

Modification of Basic Transfer function[3]

- Method II

$$Z^{-1} = \frac{Z^{-1} - \alpha}{1 - \alpha Z^{-1}}$$

$$\left(1 - \alpha Z^{-1}\right) Z^{-1} = Z^{-1} - \alpha$$

$$Z^{-1} + \alpha = Z^{-1} + \alpha \; Z^{-1} Z^{-1}$$

$$\alpha + Z^{-1} = Z^{-1}\left(1 + \alpha \; Z^{-1}\right)$$

$$Z^{-1} = \frac{\left(1 + \alpha \; Z^{-1}\right)}{\left(\alpha + Z^{-1}\right)}$$

$$F\left(Z\right) = \sum_{m=0}^{M} C(m) Z^{-m}$$

$$\left(\alpha + Z^{-1}\right) \sum_{m=0}^{M} \frac{C(m)}{\left(\alpha + Z^{-1}\right)} Z^{-m}$$

$$\left(\alpha + Z^{-1}\right) \sum_{m=0}^{M} \frac{C(m) Z^{-1}}{\left(\alpha + Z^{-1}\right)} Z^{-(m-1)}$$

$$\frac{(1 - \alpha^2) Z^{-1}}{1 - \alpha Z^{-1}} \sum_{m=0}^{M} B(m) Z^{-(m-1)}$$

$$B(m) = C(m)\left(\frac{Z^{-1}}{\alpha + Z^{-1}}\right)$$

$$B(m+1) = 0$$

$$B(m) = C(m) - \alpha B(m+1)$$

Modification of Basic Transfer function[3]

- Method III

$$F\left(Z\right)=\sum_{m=1}^{M}C(m)Z^{-m}$$

$$F\left(Z\right)=\left(\frac{\alpha+Z^{-1}}{1+\alpha Z^{-1}}\right)\sum_{m=1}^{M}C(m)\left(\frac{\alpha+Z^{-1}}{1+\alpha Z^{-1}}\right)Z^{-1}Z^{-(m-1)}$$

$$F\left(Z\right)=\left(\frac{\alpha+Z^{-1}}{1+\alpha Z^{-1}}\right)\sum_{m=1}^{M}\mathrm{b}(m)Z^{-(m-1)}$$

$$F\left(Z\right)=z^{-1}\sum_{m=1}^{M}\mathrm{b}(m)Z^{-(m-1)}$$

$$\mathrm{b}(m)=\mathrm{c(m)}\left(\frac{1+\alpha Z^{-1}}{\alpha Z+1}\right)$$

$$\alpha\,\mathrm{b}(m+1)+\mathrm{b}(m)=C(\mathrm{m})+\alpha C(\mathrm{m}-1)$$

$$\mathrm{b(M+1)}=\alpha\,\mathrm{c(M)}$$

$$\mathrm{b}(m)=C(\mathrm{m})+\alpha\left(C(\mathrm{m}-1)-\mathrm{b}(m+1)\right)$$

$$m=M,M-1,\ldots,2$$

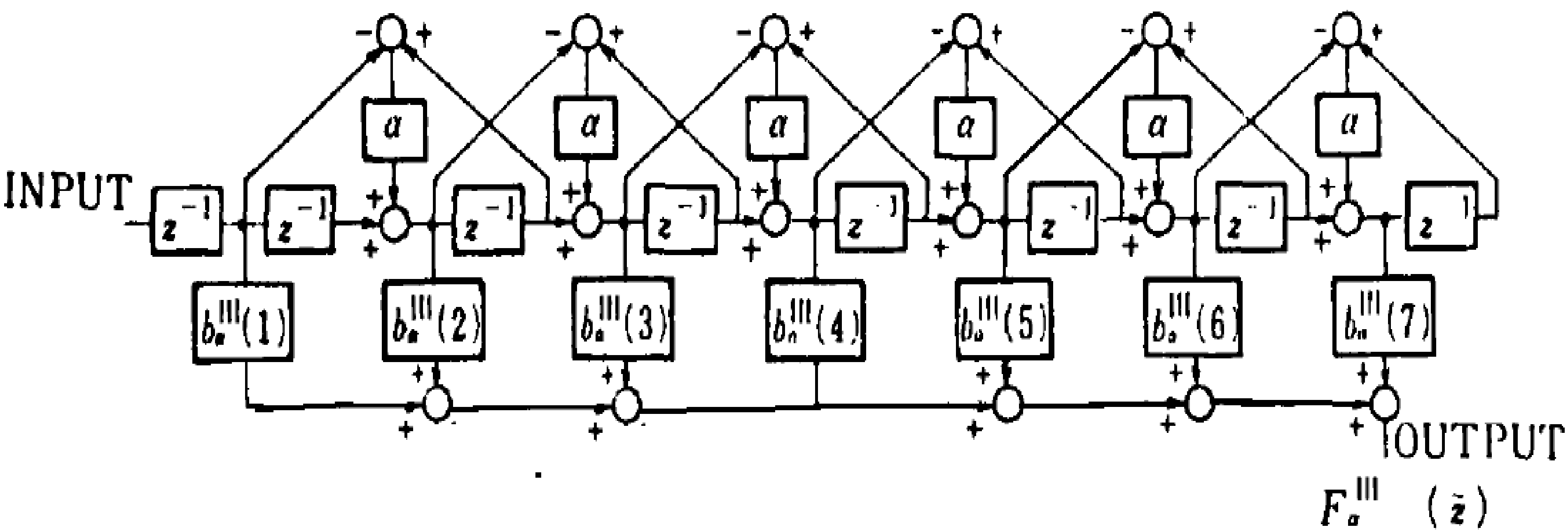
$$\mathrm{b}(1)=\frac{C_a(1)-\alpha b(2)}{(1-\alpha^2)}$$

$$\mathrm{b}(0)=C(0)-\alpha b(1)$$

Modification of Basic Transfer function[3]

- Method III

$$\mathrm{b}(m) = \mathrm{c(m)}\left(\frac{1+\alpha Z^{-1}}{\quad} \right)$$



$$F\left(Z\right) = z^{-1}\sum_{m=1}^{M} \mathrm{b}(m) Z^{-(m-1)}$$

Comparison[3]

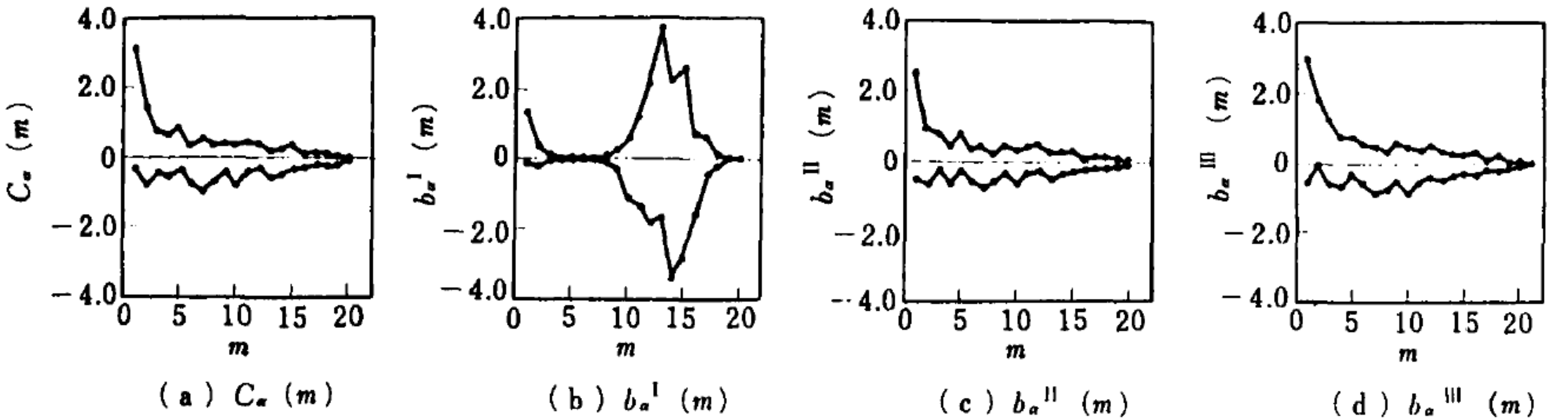


Fig. 1. Maximal and minimal values of mel cepstrum $c_\alpha(m)$ and filter coefficients $b_\alpha^K(m)$ for utterance "nambu dewa higashi no kaze" of a male speaker ($\alpha = 0.4$, $K = \mathrm{I}, \mathrm{II}, \mathrm{III}$).

Implementation[3]

- To apply Pade approximation in the range of good approximation, the transfer function of the basic filter is partitioned into the sum of b(0) and $F^n(z)$ (n = 1 . 2 ..... N) as

$$F_{\alpha}^{\mathrm{III}}(\tilde{z}) = b_{\alpha}^{\mathrm{III}}(0) + \sum_{n=1}^{N} F_{\alpha}^{\mathrm{III},n}(\tilde{z})$$

MLSA filter is given by

$$H_{\alpha}(\tilde{z}) = \exp\left(b_{\alpha}^{\mathrm{III}}(0)\right) \prod_{n=1}^{N} R_{L(n)}^{\mathrm{III},n}\left(F_{\alpha}^{\mathrm{III},n}(\tilde{z})\right)$$

Implementation[3]

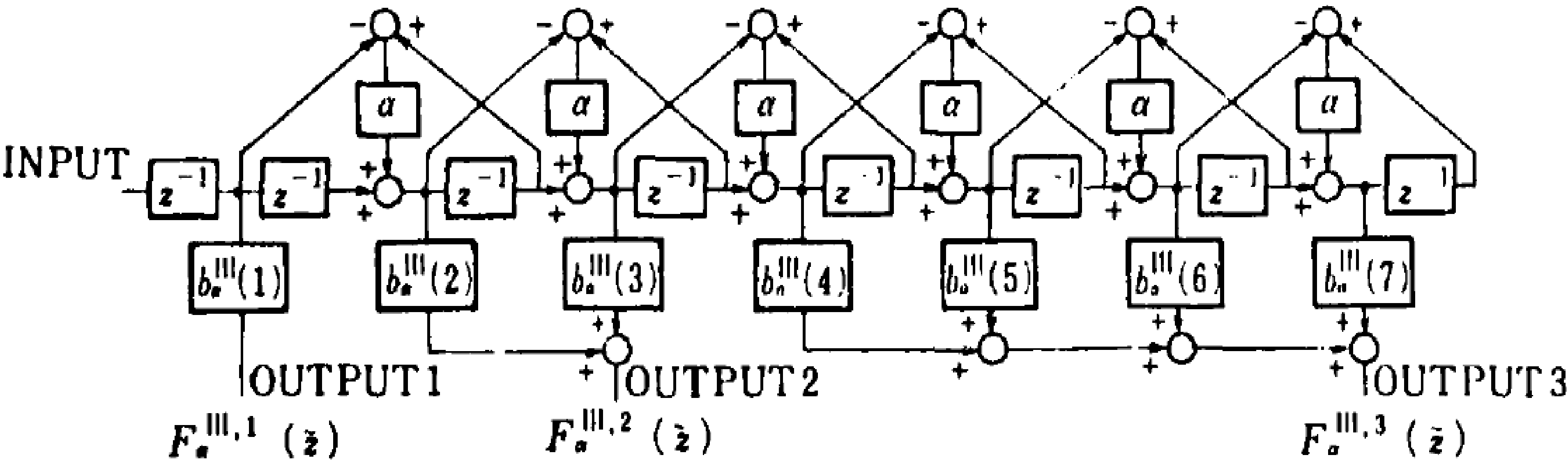
$$F_{\alpha}^{\mathrm{III},1}(\tilde{z}) = z^{-1}\, b_{\alpha}^{\mathrm{III}}(1)$$

$$F_{\alpha}^{\mathrm{III},2}(\tilde{z}) = z^{-1}\left(b_{\alpha}^{\mathrm{III}}(2)\,\tilde{z}^{-1} + b_{\alpha}^{\mathrm{III}}(3)\,\tilde{z}^{-2} \right)$$

$$F_{\alpha}^{\mathrm{III},3}(\tilde{z}) = z^{-1}\left(b_{\alpha}^{\mathrm{III}}(4)\,\tilde{z}^{-3} + b_{\alpha}^{\mathrm{III}}(5)\,\tilde{z}^{-4} + b_{\alpha}^{\mathrm{III}}(6)\,\tilde{z}^{-5} + b_{\alpha}^{\mathrm{III}}(7)\,\tilde{z}^{-6} \right)$$

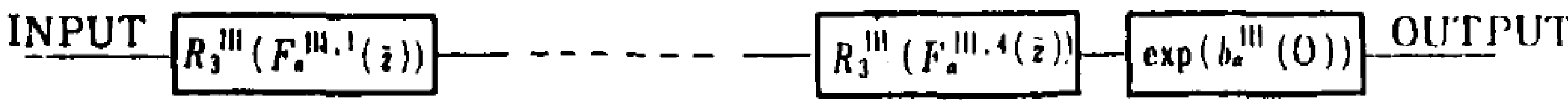
$$F_{\alpha}^{\mathrm{III},4}(\tilde{z}) = z^{-1}\left(b_{\alpha}^{\mathrm{III}}(8)\,\tilde{z}^{-7} + b_{\alpha}^{\mathrm{III}}(9)\,\tilde{z}^{-8} + \cdots + b_{\alpha}^{\mathrm{III}}(M+1)\,\tilde{z}^{-M} \right) \quad (M \leq 20)$$

Implementation[3]

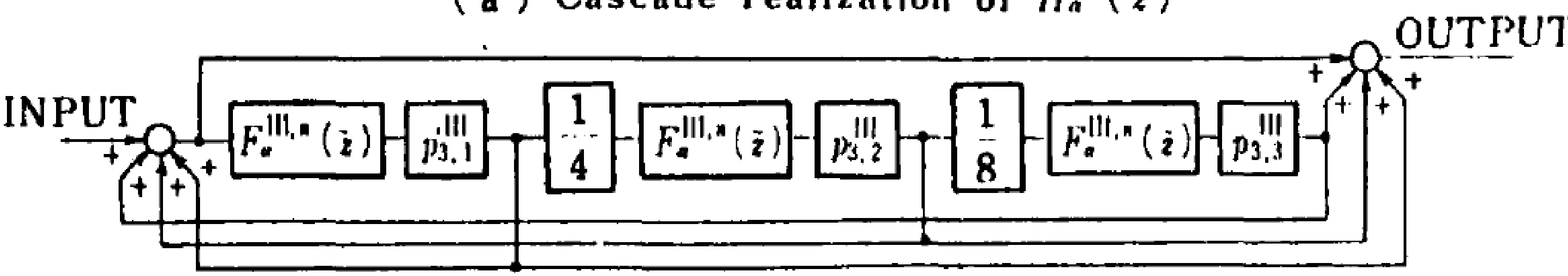


(c) $F_n^{III,n}(\tilde{z})$ $(n=1, 2, 3)$

Implementation[3]



(a) Cascade realization of H_a ($\tilde{z}$)



(b) R_3^{III} ($F_a^{III,n}$ ($\tilde{z}$)) ($n = 1, 2, 3, 4$)

References

1. Strube, Hans Werner. "Linear prediction on a warped frequency scale." The Journal of the Acoustical Society of America 68.4 (1980): 1071-1076.
2. Oppenheim, Alan V., and Donald H. Johnson. "Discrete representation of signals." Proceedings of the IEEE 60.6 (1972): 681-691.
3. Imai, Satoshi, Kazuo Sumita, and Chieko Furuichi. "Mel log spectrum approximation (MLSA) filter for speech synthesis." Electronics and Communications in Japan (Part I: Communications) 66.2 (1983): 10-18.