Speech Information Technology

Lecture 3

Parametric representation of speech signals

Information and Communications Engineering Course

Takahiro Shinozaki

2022/4/22

Lecture Plan

1. Speech communication and speech interface systems
2. Signal analysis
3. Parametric representation of speech signals
4. Basics of probability distributions
5. Principles of speech recognition and synthesis
6. Graphical model
7. Markov and Hidden Markov models
8. Weighted finite state transducer
9. Dynamic programming and Viterbi algorithm
10. Bayesian inference
11. Basics of artificial neural network
12. Neural network based speech recognition and synthesis
13. Markov decision process and dialogue systems
14. Reinforcement learning and spoken language acquisition

Motivation, outline, general topics

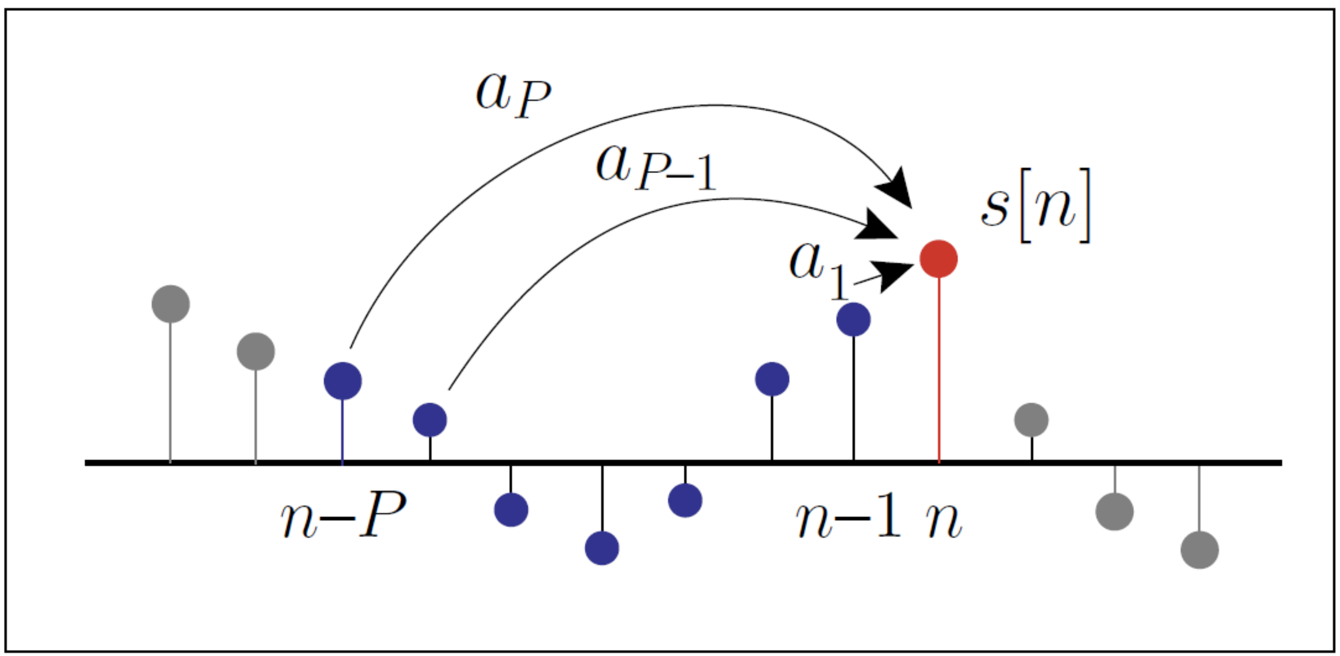
Math oriented

Linear Predictive Analysis

- Consider predicting the current sample based on past P samples by a linear regression

$$\hat{s}[n] = -\sum_{m=1}^{P} a_m s[n-m]$$

(negation is for later convenience in formulation)



LPC Coefficients

- The coefficients $\{a_1, a_2, \cdots, a_P\}$ are estimated by fitting the model to an observed speech signal
- The set of coefficients is called Linear prediction coding (LPC) coefficients, and is used as a parametric representation of the signal
- The model is fit to the signal by first defining an error measure, and then minimizing it

- Many of widely used speech coding algorithms are based on LPC
- LPC was proposed by Fumitada Itakura and Shuzo Saito in 1966

Prediction Error

- Residual signal $e[n]$ is defined as a difference of true and predicted values

$$
\begin{aligned}
e[n] &= s[n] - \hat{s}[n] \\
&= s[n] + a_1 s[n-1] + a_2 s[n-2] + \cdots + a_P s[n-P] \\
&= s[n] + \sum_{m=1}^{P} a_m s[n-m] = \sum_{m=0}^{P} a_m s[n-m] \qquad (a_0 \triangleq 1)
\end{aligned}
$$

- Residual energy is defined as a sum of squared residual signal

$$
\begin{aligned}
\sigma^2 &= \sum_{n=-\infty}^{\infty} e^2[n] \\
&= \sum_{n=-\infty}^{\infty} \left[\sum_{m=0}^{P} a_m s[n-m] \right]^2 \\
&= \sum_{m=0}^{P} \sum_{l=0}^{P} a_m a_l \sum_{n=-\infty}^{\infty} s[n-m] s[n-l]
\end{aligned}
$$

Autocorrelation

- Autocorrelation R_l at lag l is defined as:

$$R_l = \sum_{n=-\infty}^{\infty} s[n]\, s[n-l]$$

$$R_{-l} = \sum_{n=-\infty}^{\infty} s[n]\, s[n+l]$$
$$= \sum_{\acute{n}=-\infty}^{\infty} s[\acute{n}-l]\, s[\acute{n}] = R_l$$

$\acute{n} \triangleq n + l$:

- Using autocorrelation, the residual energy is rewritten as:

$$\sigma^2 = \sum_{m=0}^{P}\sum_{l=0}^{P} a_m a_l \sum_{n=-\infty}^{\infty} s[n-m]s[n-l]$$

$$= \sum_{m=0}^{P}\sum_{l=0}^{P} a_m a_l \sum_{\acute{n}=-\infty}^{\infty} s[\acute{n}]s[\acute{n}+m-l]$$

Let $\acute{n} \triangleq n - m$

$$= \sum_{m=0}^{P}\sum_{l=0}^{P} a_m\, R_{m-l}\, a_l$$

Matrix Representation of Residual Energy

$$\sigma^2 = \sum_{m=0}^{P}\sum_{l=0}^{P} a_m R_{m-l} a_l$$

$$= [a_0, a_1, \dots, a_P]\begin{bmatrix} \mathrm{R}_0 & R_1 & \cdots & R_P \\ R_1 & R_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_1 \\ R_P & \cdots & R_1 & R_0 \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_P \end{bmatrix}$$

$$= [1, \boldsymbol{a}_P^\top]\begin{bmatrix} R_0 & \boldsymbol{r}_P^\top \\ \boldsymbol{r}_P & \boldsymbol{R}_{P-1} \end{bmatrix}\begin{bmatrix} 1 \\ \boldsymbol{a}_P \end{bmatrix}$$

$$\sigma^2 = R_0 + \boldsymbol{r}_P^\top \boldsymbol{a}_P + \boldsymbol{a}_P^\top \boldsymbol{r}_P + \boldsymbol{a}_P^\top \boldsymbol{R}_{P-1} \boldsymbol{a}_P$$

R_P definition and partitioning

$$\boldsymbol{R}_P = \left[\begin{array}{c:ccc} \mathrm{R}_0 & R_1 & \cdots & R_P \\ \hdashline R_1 & R_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_1 \\ R_P & \cdots & R_1 & R_0 \end{array}\right]$$

$$\boldsymbol{a}_P = [a_1, \dots, a_P]^\top$$

$$\boldsymbol{r}_P = [R_1, \dots, R_P]^\top$$

Review of Matrix and Derivative

$$\frac{\partial}{\partial \boldsymbol{a}_P} = \left[\frac{\partial}{\partial a_1}, \frac{\partial}{\partial a_2}, \dots, \frac{\partial}{\partial a_P}\right]^\top$$

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{a}_P}(\boldsymbol{r}_P^\top \boldsymbol{a}_P) &= \frac{\partial}{\partial \boldsymbol{a}_P}(\boldsymbol{a}_P^\top \boldsymbol{r}_P) \\ &= \frac{\partial}{\partial \boldsymbol{a}_P}(R_1 a_1 + R_2 a_2 + \cdots + R_P a_P) \\ &= [R_1, R_2, \dots, R_N]^\top = \boldsymbol{r}_P,\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{a}_P}(\boldsymbol{a}_P^\top \boldsymbol{a}_P) &= \frac{\partial}{\partial \boldsymbol{a}_P}(a_1^2 + a_2^2 + \cdots + a_P^2) \\ &= [2a_1, 2a_2, \dots, 2a_P]^\top = 2\boldsymbol{a}_P\end{aligned}$$

$$\frac{\partial}{\partial \boldsymbol{a}_P}(\boldsymbol{a}_P^\top \boldsymbol{M} \boldsymbol{a}_P) = (\boldsymbol{M} + \boldsymbol{M}^T)\boldsymbol{a}_P$$

Minimization of Residual Energy

$$R_0 + \boldsymbol{r}_P^\top \boldsymbol{a}_P + \boldsymbol{a}_P^\top \boldsymbol{r}_P + \boldsymbol{a}_P^\top \boldsymbol{R}_{P-1} \boldsymbol{a}_P = \sigma^2$$

$$\frac{\partial \sigma^2}{\partial \boldsymbol{a}_P} = 2\boldsymbol{R}_{P-1}\boldsymbol{a}_P + 2\boldsymbol{r}_P = \boldsymbol{0}$$

Yule-Walker equations

$$\boldsymbol{R}_{P-1}\boldsymbol{a}_P = -\boldsymbol{r}_P \qquad \begin{bmatrix} R_0 & \cdots & R_{P-1} \\ \vdots & \ddots & \vdots \\ R_{P-1} & \cdots & R_0 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = -\begin{bmatrix} R_1 \\ \vdots \\ R_P \end{bmatrix}$$

In general, the computational cost to solve a system of linear equations is $O(P^3)$.
(e.g. Gaussian elimination)

Toeplitz Matrix

- Toeplitz matrix is a matrix in which each descending diagonal from left to right is constant

$$\begin{bmatrix} a & b & c & d \\ e & a & \ddots & c \\ f & \ddots & \ddots & b \\ g & f & e & a \end{bmatrix}$$ *Toeplitz* matrix

- When a system of linear equations has a Toeplitz matrix, Levinson-Durbin Algorithm is applicable. It is an efficient recursive algorithm that runs in $O(P^2)$

$$\begin{bmatrix} R_0 & \cdots & R_{P-1} \\ \vdots & \ddots & \vdots \\ R_{P-1} & \cdots & R_0 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = - \begin{bmatrix} R_1 \\ \vdots \\ R_P \end{bmatrix}$$

Solves order P system using the result of order P-1 system

$$\begin{bmatrix} R_0 & R_1 & R_2 \\ R_1 & R_0 & R_1 \\ R_2 & R_1 & R_0 \end{bmatrix} \begin{bmatrix} a_{3,1} \\ a_{3,2} \\ a_{3,3} \end{bmatrix} = - \begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix}$$

$$\begin{bmatrix} R_0 & R_1 & R_2 & R_3 \\ R_1 & R_0 & R_1 & R_2 \\ R_2 & R_1 & R_0 & R_1 \\ R_3 & R_2 & R_1 & R_0 \end{bmatrix} \begin{bmatrix} a_{4,1} \\ a_{4,2} \\ a_{4,3} \\ a_{4,4} \end{bmatrix} = - \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{bmatrix}$$

Levinson-Durbin Algorithm

$Initialization$: $\sigma_0^2 = R_0$

$Iteration$: For $i = 1, \ldots, P$, calculate

$$k_i = -\frac{R_i + \boxed{\sum_{m=1}^{i-1} a_{i-1,m} R_{i-m}}}{\sigma_{i-1}^2}$$

$$\boxed{a_{i,m} = a_{i-1,m} + k_i a_{i-1,i-m}}, \ m = 1, \ldots, i-1$$

$$a_{i,i} = k_i$$

$$\sigma_i^2 = \left(1 - k_i^2\right)\sigma_{i-1}^2$$

$Solution$: $\sigma^2 = \sigma_P^2$ and $a_m = a_{P,m}, \ m = 1, \ldots, P$

Regards 0 or skip when i = 1

Z-Transform of LP Analysis

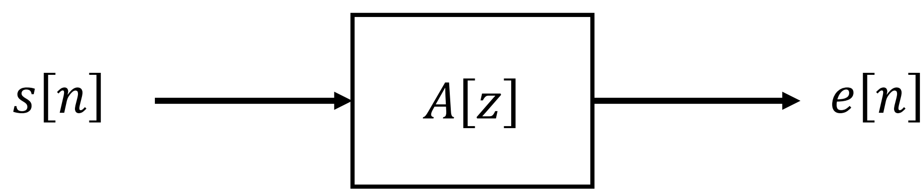
- Residual signal $e[n]$

$$
\begin{aligned}
e[n] &= s[n] - \hat{s}[n] \\
&= s[n] + a_1 s[n-1] + a_2 s[n-2] + \cdots + a_P s[n-P]
\end{aligned}
$$

- Z-transform

$$E(z) = S(z)(1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_P z^{-P}) = S(z)A(z)$$

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_P z^{-P}$$



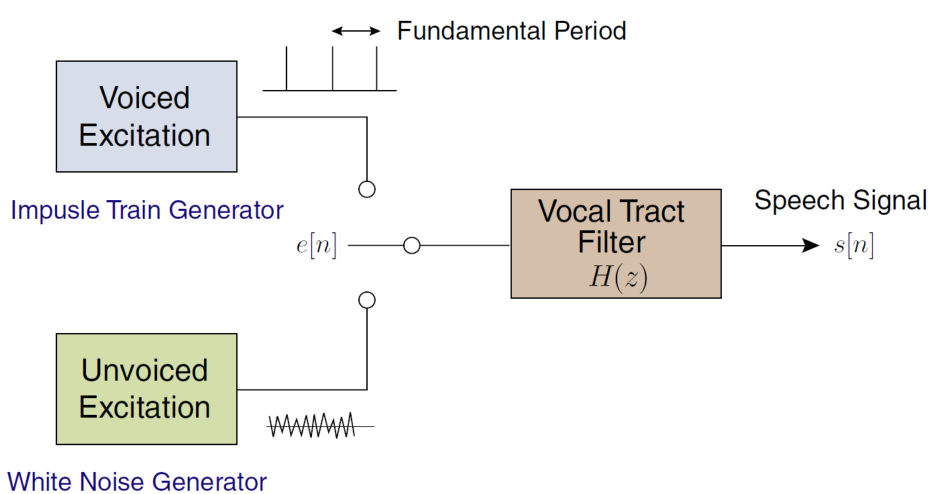
Speech Synthesis Filter by LPC Coefficients

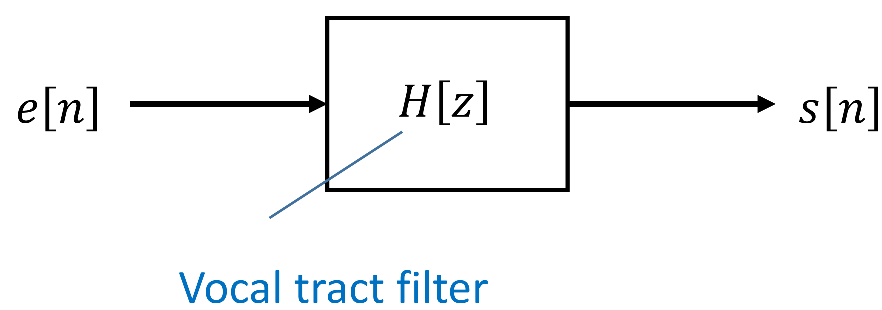
$$S(z)A(z) = E(z)$$

$$S(z) = \frac{1}{A(z)}E(z) = H(z)E(z) \qquad H(z) \triangleq \frac{1}{A(z)}$$

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_P z^{-P}}$$

All pole model

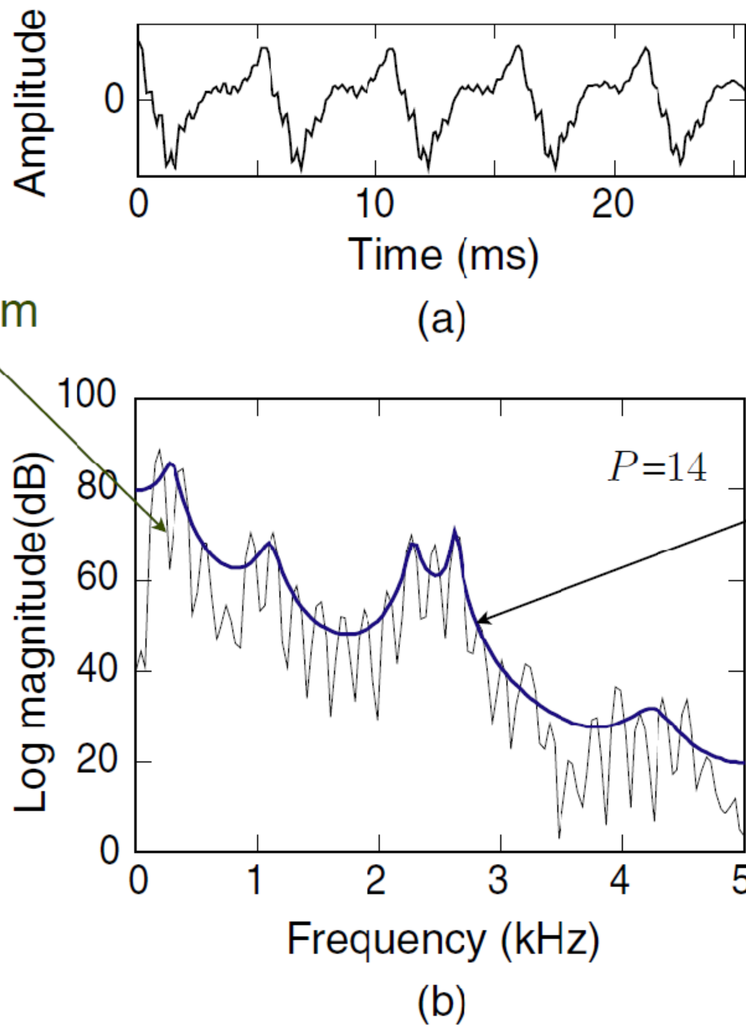




$A(z)$ is called as an LPC inverse filter

LPC Spectrum

Fourier transform of $s[n]$



Waveform $s[n]$

Frequency response of vocal tract filter $H(z)$

$$z = exp(j\omega) = exp\left(j\frac{2\pi f}{f_s}\right)$$

$$f = \frac{\omega}{2\pi} f_s$$

f : frequency

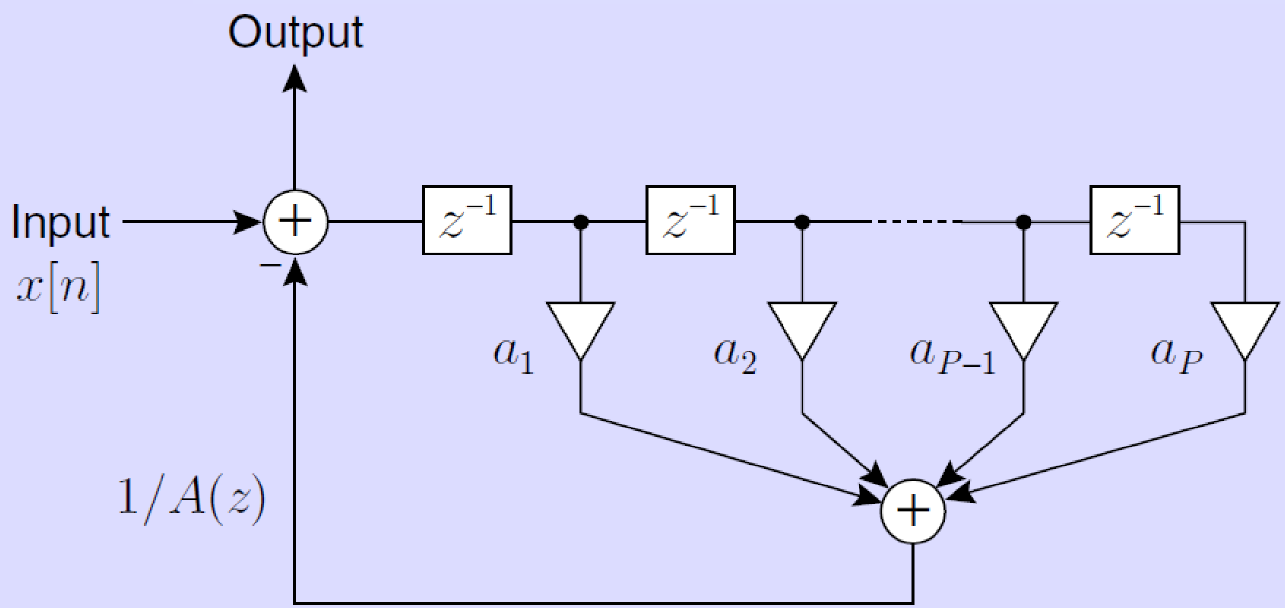
f_s : Sampling frequency

ω : Normalized angular frequency

Digital Filter Realization of All Pole Model

$$H_0(z) = \frac{1}{A(z)} = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_P z^{-P}}$$

$$y[n] = x[n] - a_1 y[n-1] - a_2 y[n-2] - \cdots - a_P y[n-P]$$



Exercise

Suppose $P = 2$, $R_0 = 10, R_1 = 5, R_2 = -2$.
Using the Levinson-Durbin Algorithm, answer the following questions

3-1)

Obtain k_1

3-2)

Obtain a_2

PARCOR Analysis

- Problem of LPC
  - When LPC is used to analyze and re-synthesize speech sound as a compression method for digital communication, high accuracy is required to store the coefficients $\{a_1, a_2, \cdots, a_P\}$
  - If the quantization error is large to represent the coefficients, the re-synthesis becomes unstable
- PARCORE
  - Robust for the quantization error
  - PARCORE and LPC coefficients are mutually convertible (cf. lpc2par and par2lpc commands in SPTK. See appendix for more details)

- Proposed by Fumitada Itakura and Shuzo Saito in 1969

Forward and Backward Prediction Errors

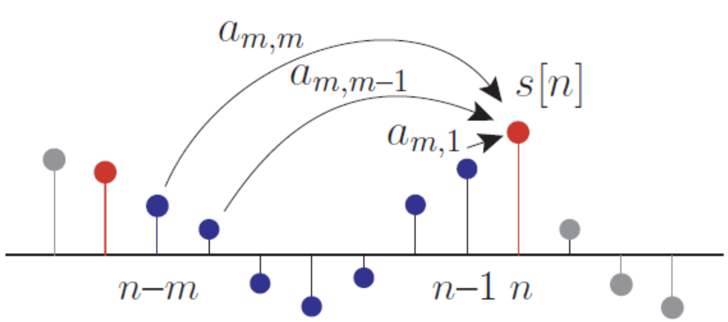
$e_m^f[n]$: Forward prediction error, $e_m^b[n]$: Backward prediction error

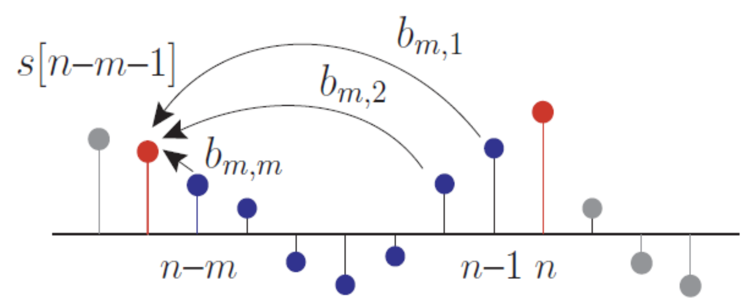
For $m = 0$: $\quad e_0^f[n] = s[n], e_0^b[n] = s[n-1]$

For $m = 1, 2, \cdots, P-1$:

$$e_m^f[n] = s[n] + a_{m,1}s[n-1] + a_{m,2}s[n-2] + \cdots + a_{m,m}s[n-m]$$

$$e_m^b[n] = b_{m,1}s[n-1] + b_{m,2}s[n-2] + \cdots + b_{m,m}s[n-m] + s[n-m-1]$$





$a_{1,1}$
$a_{2,1}, a_{2,2}$
$\vdots$
$a_{P,1}, a_{P,2}, \cdots, a_{P,P}$

$b_{1,1}$
$b_{2,1}, b_{2,2}$
$\vdots$
$b_{P,1}, b_{P,2}, \cdots, b_{P,P}$

PARCORE Coefficients

- PARCORE coefficient k_m between $s[n]$ and $s[n-m]$ is the correlation coefficient between $e_{m-1}^{f}[n]$ and $e_{m-1}^{b}[n]$, for $m = 1,2,\cdots,P$

$$k_m = \frac{\sum_{n=-\infty}^{\infty} e_{m-1}^{f}[n] e_{m-1}^{b}[n]}{\sqrt{\sum_{n=-\infty}^{\infty}\left(e_{m-1}^{f}[n]\right)^2}\sqrt{\sum_{n=-\infty}^{\infty}\left(e_{m-1}^{b}[n]\right)^2}}$$

PARCOR and LPC Coefficients

- LPC coefficients of order P

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_P z^{-P}$$

- Consider a set of LP models of order m ($m = 1,2, \cdots, P$)

$$A_m(z) = 1 + a_{m,1} z^{-1} + a_{m,2} z^{-2} + \cdots + a_{m,m} z^{-m}$$

where $A_P(z) = A(z), a_{P,i} = a_i, i = 1, 2, \dots, P$

- All the coefficients $\{a_{m-1,k}\}$ are obtained recursively from $\{a_{P,i}\}$

$$a_{m-1,k} = \frac{a_{m,k} - a_{m,m} a_{m,m-k}}{1 - a_{m,m}^2} \qquad \begin{array}{l} m = P, P-1, \cdots, 2 \\ k = 1,2, \cdots, m-1 \end{array}$$

- PARCORE coefficients k_m is equal to $a_{m,m}$

$$k_m = a_{m,m}, m = 1,2, \cdots, P$$

Durbin's Recursion and PARCOR

- PARCOR coefficients are equal to the variables k_i in the Durbin's recursion used to obtain LPC coefficients

$$Initialization: \sigma_0^2 = R_0$$

$$Iteration: \text{For } i = 1, \ldots, P, \text{calculate}$$

$$k_i = -\frac{R_i + \sum_{m=1}^{i-1} a_{i-1,m} R_{i-m}}{\sigma_{i-1}^2}$$

$$a_{i,m} = a_{i-1,m} + k_i a_{i-1,i-m}, \ \ m = 1, \ldots, i-1$$

$$a_{i,i} = k_i$$

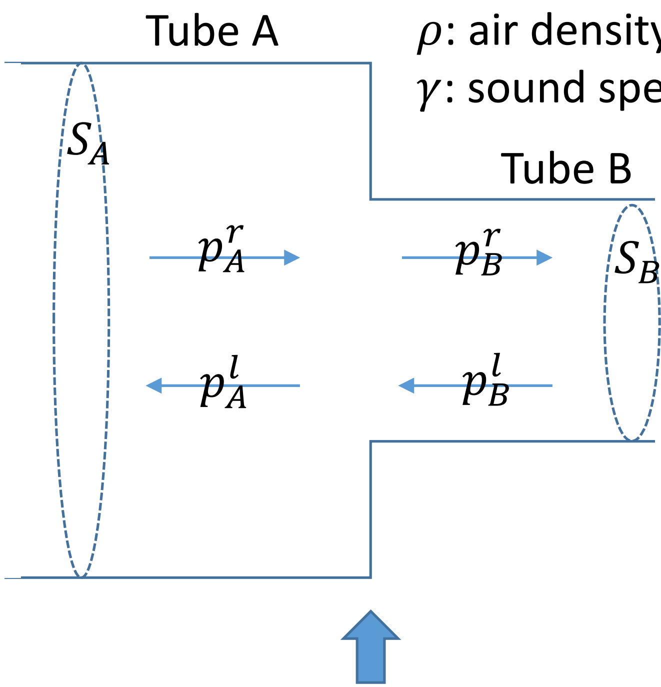
$$\sigma_i^2 = \left(1 - k_i^2\right)\sigma_{i-1}^2$$

$$Solution: \ \sigma^2 = \sigma_P^2 \text{ and } a_m = a_{P,m}, \ \ m = 1, \ldots, P$$

Acoustic Tube Model

Let two tubes A and B are concatenated, where their cross section area are S_A and S_B.
Let p_A and p_B be pressure deviations from the mean pressure, and u_A and u_B be volume velocities in tube A and B, respectively.
Let p_A^l and p_A^r be left and rightward traveling pressure waves in tube A. Similarly, let p_B^l and p_B^r be left and right waves in tube B.

ρ: air density
γ: sound speed



Junction of the two tubes

We assume wave length is longer than the tube radius.
According to laws of physics, the following equations hold at the junction:

$$p_A = p_B$$
$$u_A = u_B$$

$$p_A = p_A^r + p_A^l$$
$$p_B = p_B^r + p_B^l$$

$$u_A = Y_A\left(p_A^l - p_A^r\right) \qquad Y_A = \frac{S_A}{\rho\gamma}$$
$$u_B = Y_B\left(p_B^l - p_B^r\right) \qquad Y_B = \frac{S_B}{\rho\gamma}$$

Reflection of Waves

- By solving the system for p_B^r and p_A^l, the following relation is obtained, where R is reflection parameter

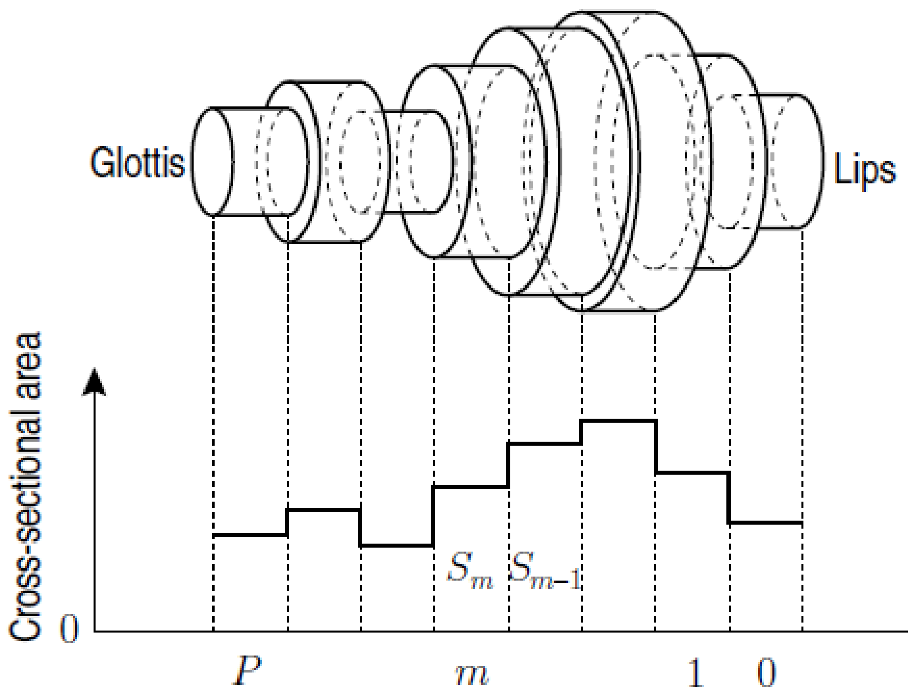
$$p_A^l = Rp_A^r + (1 - R)p_B^l$$

$$p_B^r = (1 + R)p_A^r - Rp_B^l$$

$$R = \frac{Y_A - Y_B}{Y_A + Y_B} = \frac{S_A - S_B}{S_A + S_B}$$

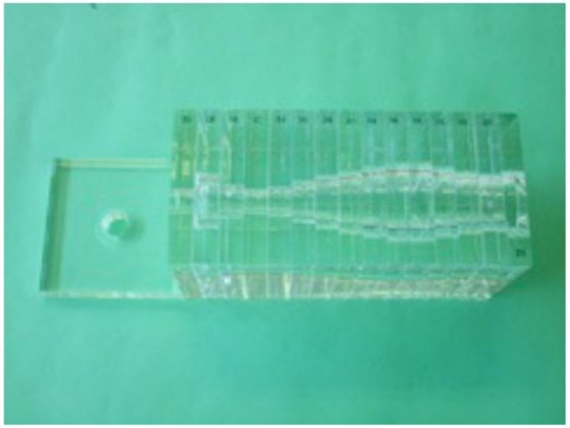
- Observation with the reflection
  - When S_A and S_B is equal, no reflection occurs
  - Let $p_B^l = 0$
    - When S_B is very small compared to S_A, p_A^r is completely reflected and p_A^l becomes equal to p_A^r. It is an open-end reflection for pressure wave, since the pressure can take arbitral value
    - When S_B is very large compared to S_A, p_A^r is completely reflected and p_A^l becomes equal to $-p_A^r$. It is an closed-end reflection for pressure wave, since the pressure in B hardly change due to the large volume

Tube based Vocal Tract Model



$$R_m = \frac{Y_m - Y_{m-1}}{Y_m + Y_{m-1}} = \frac{S_m - S_{m-1}}{S_m + S_{m-1}}$$

$$m = P, P-1, \cdots, 1$$



The PARCORE coefficients k_m are identical to $-R_m$, where $|R_m| < 1$

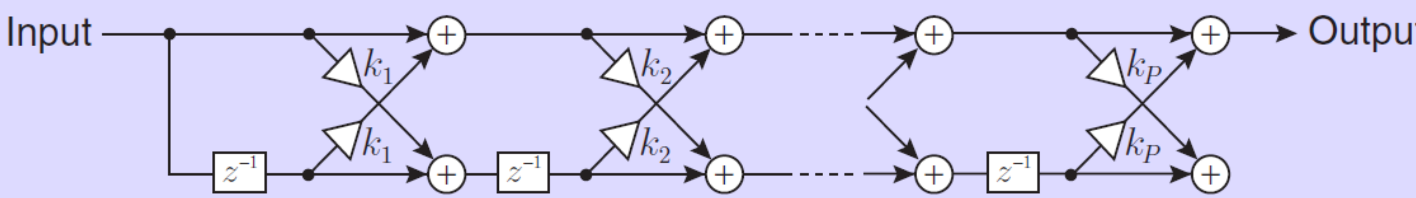
$$k_m = -R_m$$

[J. D. Markel and A. H. Gray Jr., Linear Prediction of Speech, Springer, 1976]

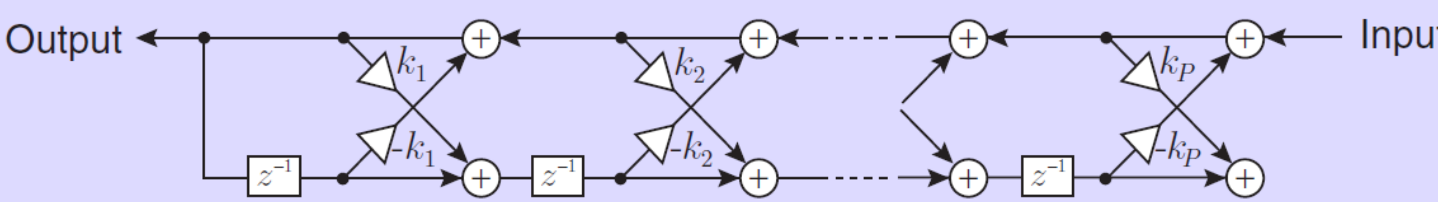
PARCOR Synthesis Filter

- *Lattice* form realization of the all-pole transfer function
- Stability for $1/A(z)$
  - Necessary and sufficient condition

    $|k_m| < 1$ for $1 \le m \le P$



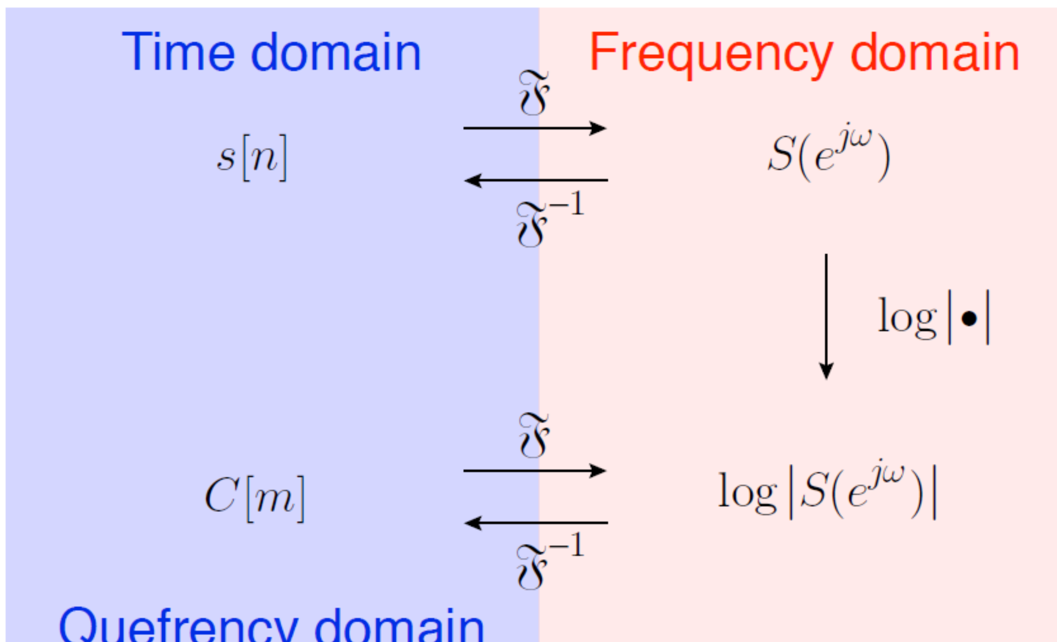
(a) $A(z)$



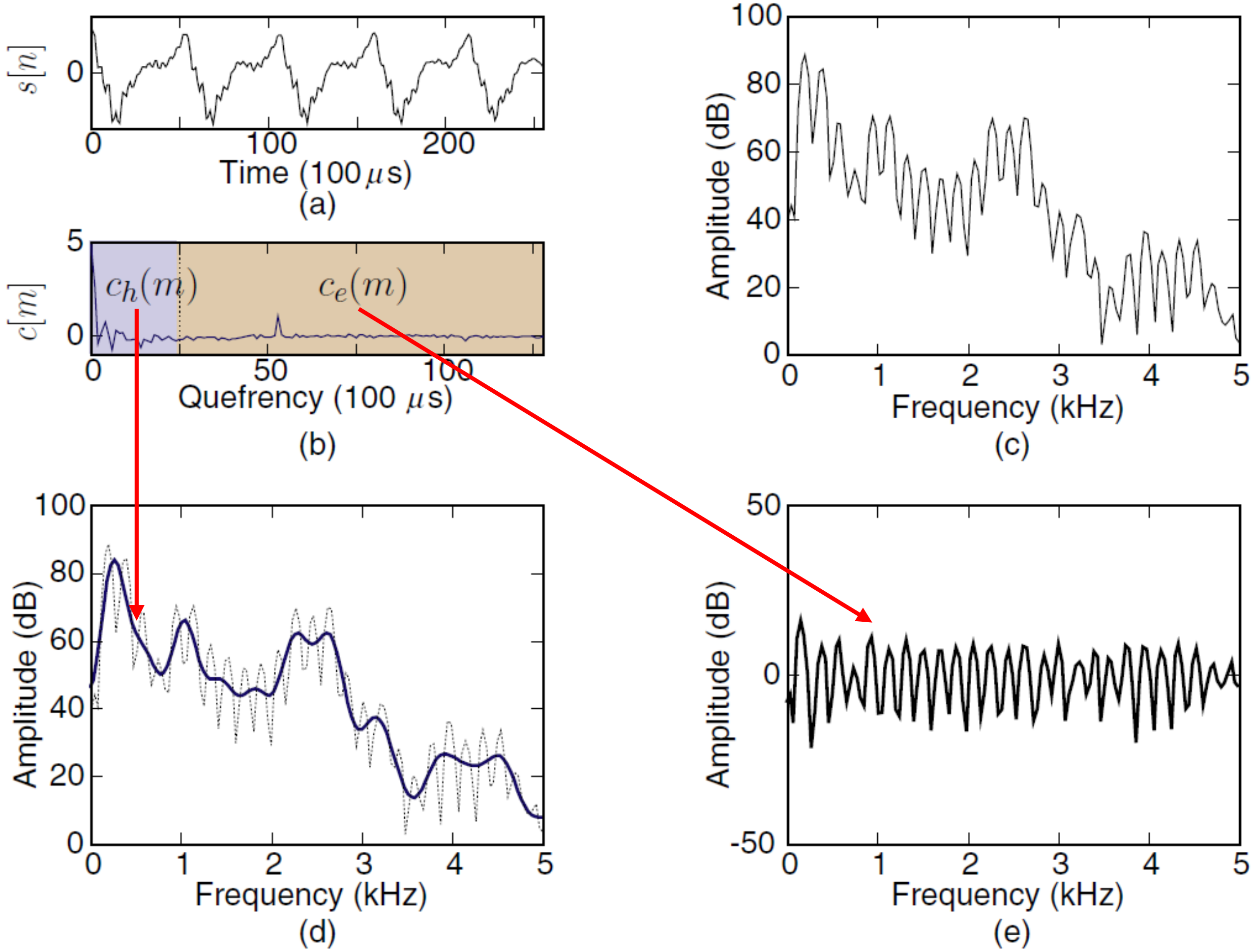
(b) $1/A(z)$

Cepstrum

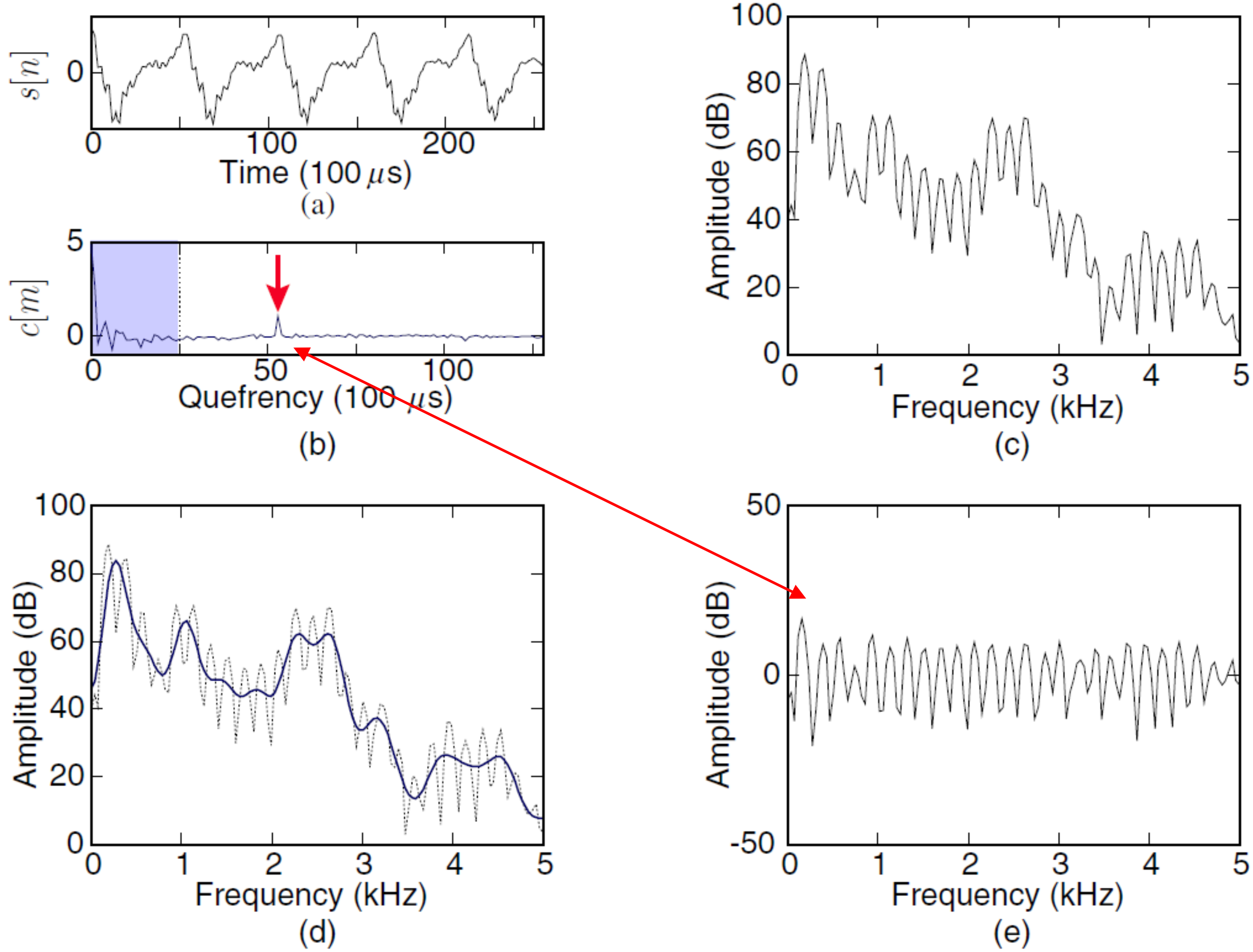
- Terminology
  - *Cepstrum*: anagram of spectrum
  - *Quefrency*: anagram of frequency
- Definition



Example of Cepstral Analysis



F0 Estimation Based on Cepstral Analysis

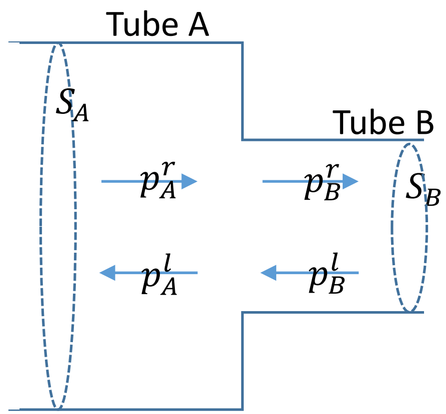


Reciprocal of the peak position in Qefrency corresponds to the Fundamental frequency

e.g. Quefrency 0.005 (sec) → F0 200 (Hz)

Exercise

Consider the following acoustic tube.



3-3)

When $S_A = 2$, $S_B = 1$, $p_A^r = 4$, $p_B^l = 1$, obtain p_A^l

Exercise (How to submit)

- Deadline: 2023/4/24 13:00
- Upload to : T2SCHOLA

Appendix

Derivation Process of the Reflection

$$p_A = p_B \qquad p_A = p_A^r + p_A^l \qquad u_A = Y_A(p_A^l - p_A^r)$$
$$u_A = u_B \qquad p_B = p_B^r + p_B^l \qquad u_B = Y_B(p_B^l - p_B^r)$$

We want to solve the system for p_A^l and p_B^r.
For simplicity of notation, let's put

$$x_1 = p_A^r, x_2 = p_B^l, y_1 = p_A^l, y_2 = p_B^r, A = Y_A, B = Y_B$$

$$\Rightarrow \quad \begin{aligned} x_1 + y_1 &= y_2 + x_2 \\ A(y_1 - x_1) &= B(x_2 - y_2) \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} y_1 &= \frac{(A-B)x_1 + 2Bx_2}{A+B} \\ y_2 &= \frac{2Ax_1 - (A-B)x_2}{A+B} \end{aligned}$$

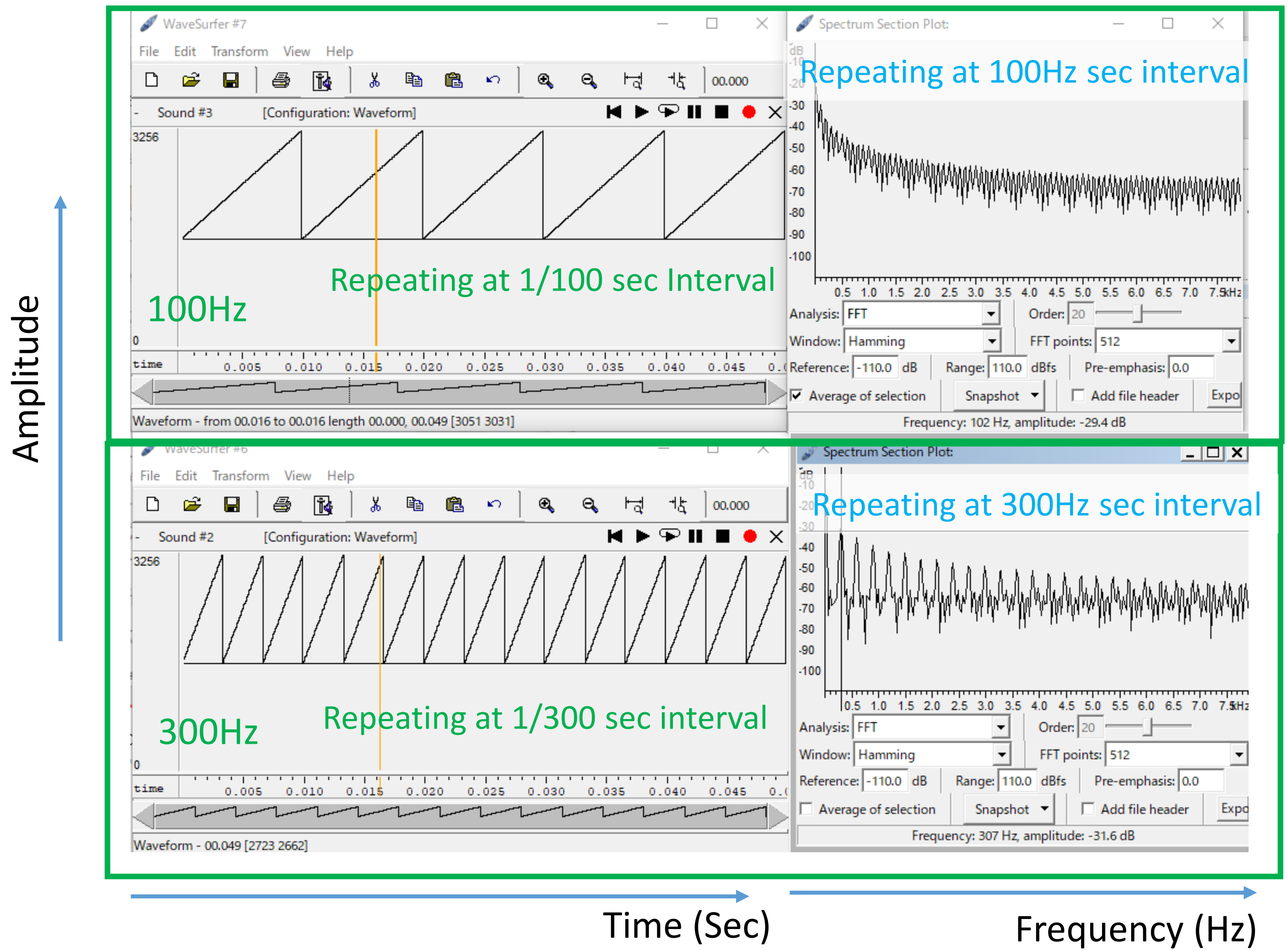
Let $R = \frac{A-B}{A+B}$

$$1 + R = \frac{2A}{A+B}$$

$$1 - R = \frac{2B}{A+B}$$

$$\Rightarrow \quad \begin{aligned} y_1 &= Rx_1 + (1-R)x_2 \\ y_2 &= (1+R)x_1 - Rx_2 \end{aligned}$$

C.f. Spectrum of Sawtooth Wave



Related Documents

- 板倉 文忠、「統計的手法による音声分析合成系に関する研究」、名古屋大学博士学位論文、1972
  PDF copy is available at 名古屋大学学術機関リポジトリ
- 板倉 文忠、「私の本棚紹介」、IEICE 情報・システムソサイエティ誌、第6巻4号、2002
  Behind-the-scenes story is introduced
  https://www.ieice.org/iss/jpn/Publications/society_mag/pdf/Vol6No4.pdf
- Noboru Sugamura and Fumitada Itakura, "Speech analysis and synthesis methods developed at ECL in NTT --From LPC to LSP—", Speech Communication, Volume 5, Issue 2, June 1986, pp. 199-215
  https://www.sciencedirect.com/science/article/pii/0167639386900087
  (Tokyo Tech have a license and the PDF is available for free inside University network. From your home, you can access the network by using VPN from Tokyo Tech Portal)
- Stefan Bilbao, "Wave and Scattering Methods for the Numerical Integration of Partial Differential Equation," Ph.D. Thesis, 2001
  https://ccrma.stanford.edu/~bilbao/master/goodcopy.html
  https://ccrma.stanford.edu/papers/wave-and-scattering-methods-numerical-integration-of-partial-differential-equation

Related Software

- Speech Signal Processing Toolkit (SPTK)
  http://sp-tk.sourceforge.net/