

()

. 『 - 』. 1996, 1, 104-125.

가 가 ,

EGG, EMA, EPGG

가

가

가

formant
Animated Vocal Tract Profiles

I.

가

Video-CD

, TV 가

가

() ,

(intelligibility)

(naturalness)

가 , 가

() (masking)

가

MPEG Video-CD HDTV (Gersho, 1994).

가

가

가

가

64Kbps(8KHz × 8 bit) (

VSELP) 8Kbps

4.8Kbps 가

(

4.8Kbps naturalness가).

(articulatory movement estimation) (articulatory speech

synthesizer) (Schroeter and Sondhi, 1994).

(feedback)

가 ,

가

II.

1. :

가

(vocal tract)

s(1967)

Schroeder (1967) Mermelstein

Paige and Zue (1970)

Atal and Hanauer (1971)

Wakita (1973)

가

가

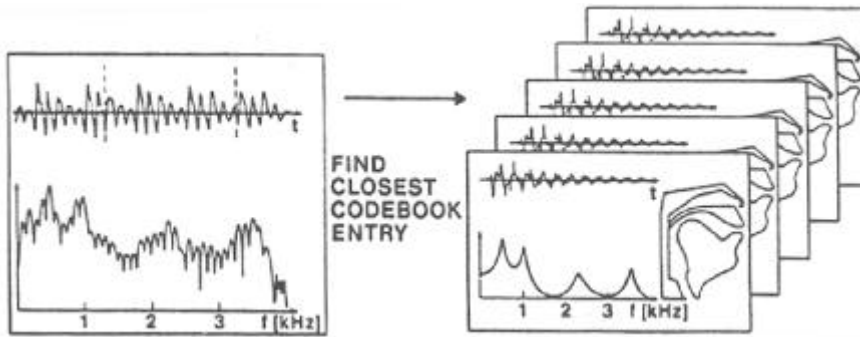
가.

가

< -1> Schroeter and Sondhi (1994)

가

가



< - 1 >

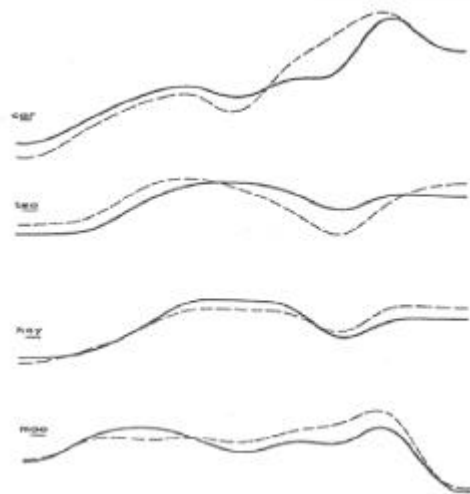
(: Schroeter and Sondhi, 1994)

all-pole model , pulse
 가 가
 가 가 가
 가

(Wakita, 1973). < -2> Crichton and Fallside (1974)가

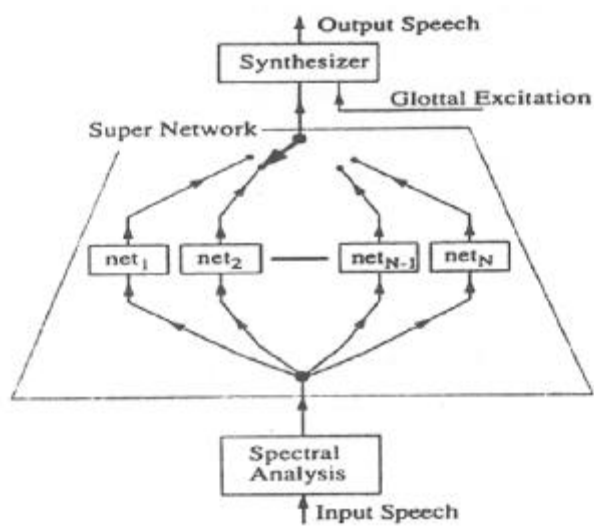
Acoustic-Articulatory Mapping

가
 (Rahim et al., 1993; Guenther, 1994). < -3> Rahim et
 al.(1993) 18



< -2>

(: Crichton and Fallside, 1974)



< -3>

(: Rahim et al., 1994)

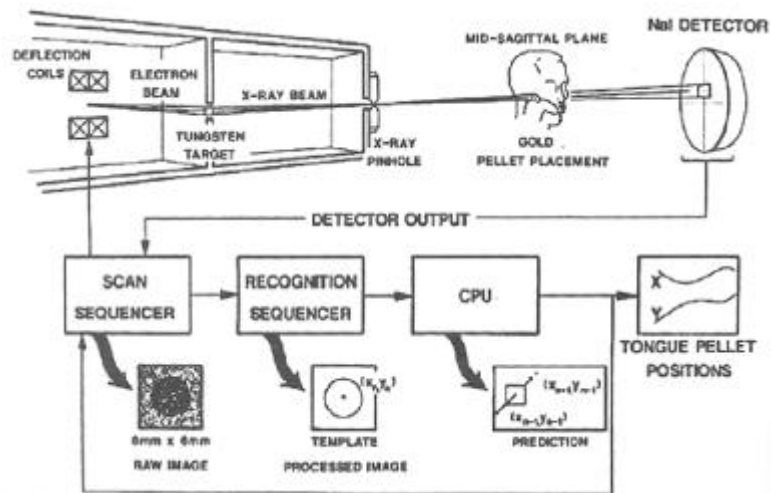
가 가
 가 (transition state)
 가 (anti-resonance)
 가 . Atal et al.(1978)

2 .

가 가
 가
 가 가
 EPG(electro-palatograph) EMA(electromagnetic articulograph)

가. X-Ray X-Ray Microbeam

X-ray video
 가
 ()
 microbeam . < -4> Wisconsin X-ray
 microbeam MRI



< -4> Microbeam X-ray

(: <http://www.biostat.wisc.edu/ubeam/images/shematic.gif>)

. EPG (Electro-Palatograph)

가

. < -5>

EPG

(dorsum)

. < -5>

Reading University EPG 3

62

< -6>

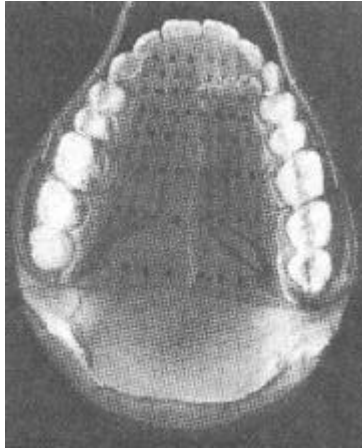
/s/

EPG

(Javkin et al., 1993).

가

가



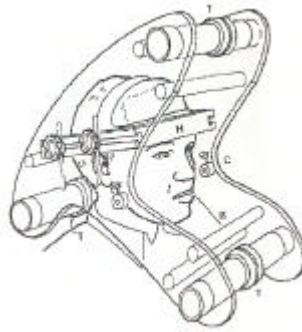
< -5> EPG model 3
 (: <http://midwiche.reading.ac.uk/research/speechlab/epg>)



< -6> /s/ EPG
 (: Javkin et al., 1993)

. EMA (Electromagnetic Articulograph)

EPG 가
 EMA가 . < -7> 3 EMA
 3 5
 가 (induced)
 가



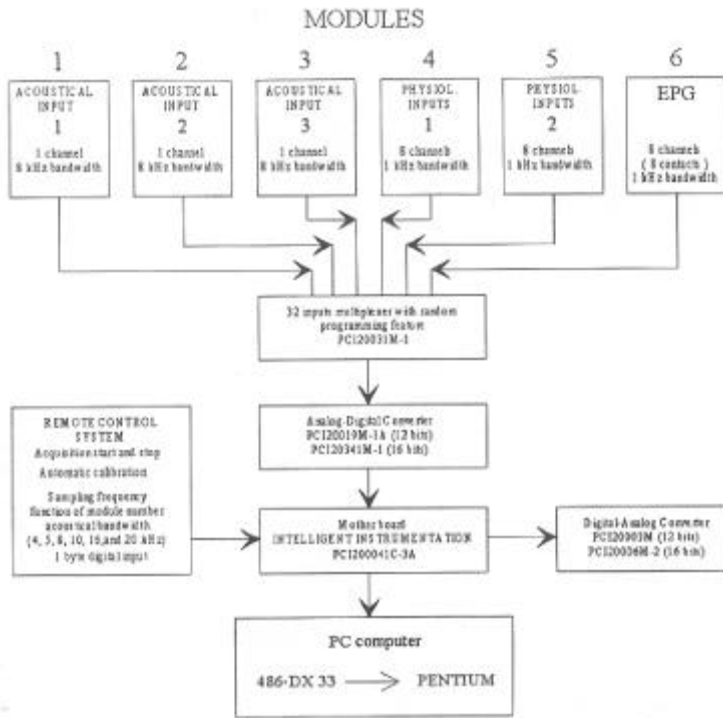
< -7> 3 EMA
 (: Perkell et al., 1992)

가

- EGG (electro-glottalgraph)
- Airflow transducer Pneumotachograph
- EPG (electro-palatograph)

1982).

(,



< - 8 >

Physiologia

(: <http://www.lpl.univ-aix.fr/valorisation/physiologia>)

III. LPC-Formant Based Vocal Tract Profile Graphics

LPC (Linear Prediction Coefficient)-Formant Based Vocal Tract Profile Estimation
(1992)

opening, rounding degree

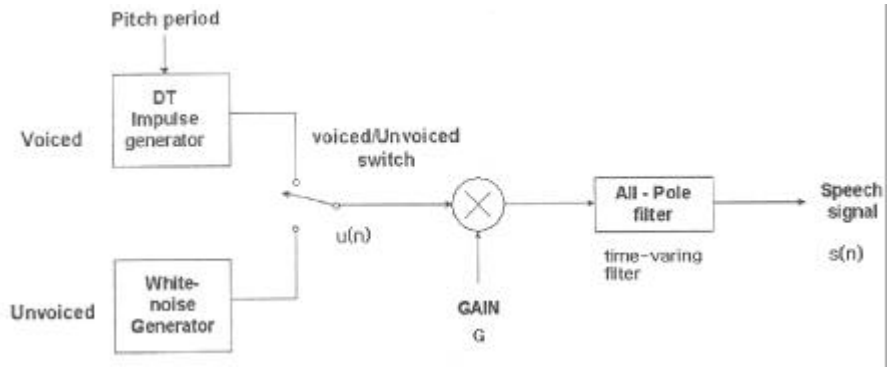
animated trajectory

1. LPC

가 가

LPC

. LPC < -9> .



< -9>

< -9> G, u(n) . u(n)
 가 , random (white noise)
 . (time-varying digital filter)

, 가 . s(n)
 가 . < -9> all-pole
 (1) (, G: , α_i : , p:).

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{i=1}^p \alpha_i z^{-i}} \quad (1)$$

(1) s(n) AR

, α_i G , (a_i) (minimum
 mean squared error; min)가 , α_i G .

2. LPC

LPC (linear prediction method)

(auto-regressive; AR)

3 . Fo r m a n t

Formant (, ,)

17cm , 3KHz 3
 4 , 5KHz 4 5 (formant)
 가 , 가 가
 가 , 가 () 가 ,
 가 .
 LPC
 가 , non-formant peak가
 (McCandless, 1974). LPC
 가 가 (merging)
 1, 2 .

4 .

가.

(vocal tract profile) ,
 (vocal tract) .
 가
 , 가 1mm
 (resolution) .

(1)

(2) 15

(3) X-

(4)

가 가 , 가

가 가

가 가 ,

가 , 15

LPC

15 pipe

17cm 17cm

section

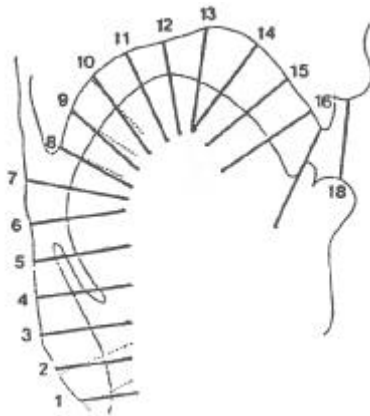
, X- Ladefoged et al.(1978) PARAFAC

17 section 5 10

grid-line

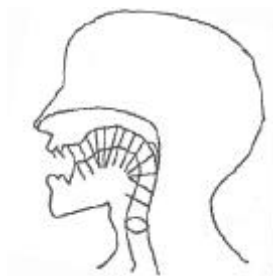
grid-line . < - 10> Ladefoged

et al.(1978) 17 section



< -10> Harshman X-data

() 11 , 2 section, 15 section
 가
 가
 가
 , aspect
 ratio가 2.1 : 1 가 2 : 1 . < -11>



< -11>

15 section

LPC $A_n(z)$ (2)

$$A_n(z) = \frac{\text{forward volume velocity}}{\text{propagated volume velocity}} \quad (2)$$

, n: section

, section n ϕ_n

$$\phi_n = \phi_{n-1} \frac{1 + C_n}{1 - C_n} \quad (3)$$

C_n inverse filter ϕ_n section ϕ_{n-1} Y

(4) (4) PARAFAC

$$X_1 = c_1 F_2 + c_2 F_2 F_3 + c_3 (F_1 / F_2) + c_4 \quad (4)$$

, $c_1 = 0.300 \times 10^{-3}$, $c_2 = -0.343 \times 10^{-6}$,
 $c_3 = 4.143$, $c_4 = -0.174$

X_1 X_3

$$X_2 = \frac{X_1 + X_3}{2} \quad (5)$$

rounding degree, height formant

(mapping)

가
 가
 (X_p, Y_p) , (X_a, Y_a)
 X_p, Y_p (6)

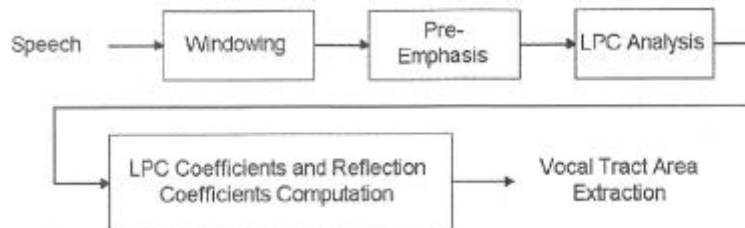
$$X_p = X_a + \frac{MX_a + NX_d}{M + N}, \quad Y_p = Y_a + \frac{MY_a + NY_d}{M + N} \quad (6)$$

$$M = \frac{1}{[(X_a - X_d)^2 + (Y_a - Y_d)^2]^{1/2}}$$

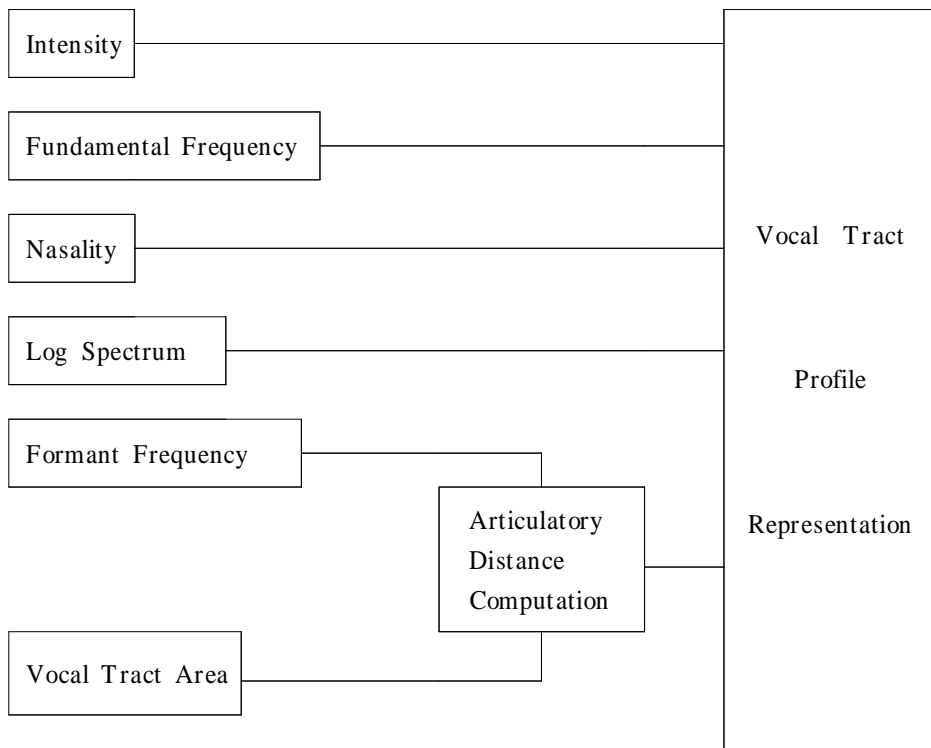
$$N = 1 - M$$

X_d, Y_d :

< -12> < -13>



< -12>



< - 13 >

IV.

4 가 5 / , , , , /

< - 14 >

, 4 가 / , , , , /

, Harshman X-

< - 15 > X- , 가

< - 16 >

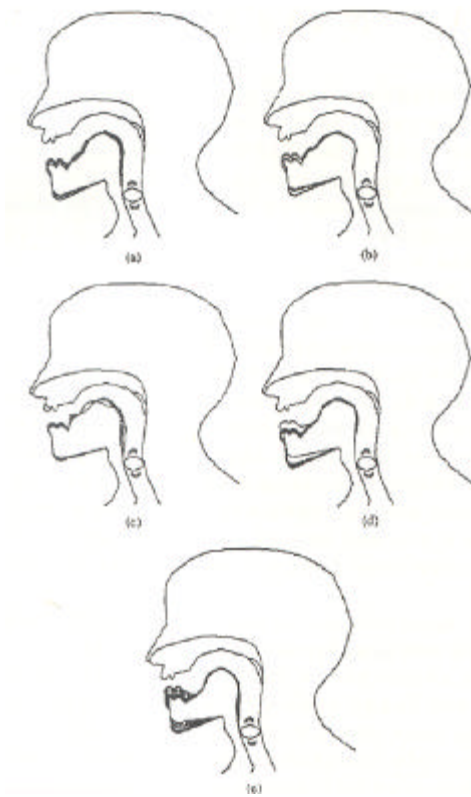
(intensity), (pitch), (nasality), (spectrum)

가

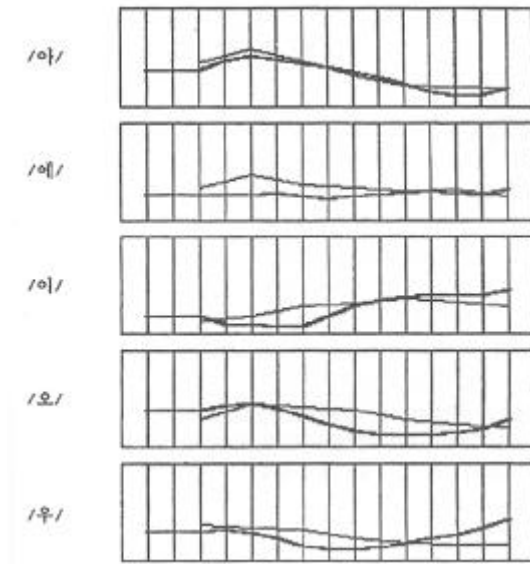
< -16> / / / / 가 / / 가 / / 가 (Park et al., 1994).

LPC-Formant Based Vocal Tract Profile Graphics LPC opening rounding degree

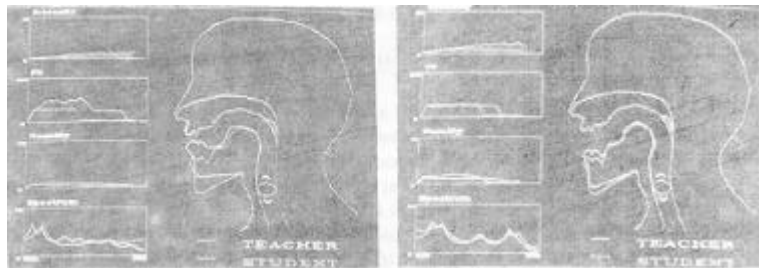
LPC glottal source pseudo-glottal waveform glottal source vocal tract filter pitch-synchronous two channel analysis



< -14> / /, / /, / /, / /, / / 5



< -15> X-



< -16> (/ / , / /)

V.

, . 가

EGG, EMA, EPGG

가 가

가 ,

가

가 가

(1992). 『

』, 41, 209-216.

(1982). 『

』, 5, 38-44.

Atal, B. S. and Hanauer, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America*, 50, 637-655.

Atal, B. S., Chang, J. J., Mathews, M. V., and Tukey, J. W. (1978). Inversion of articulatory to acoustic transformation in the vocal tract by a computer sorting technique. *Journal of the Acoustical Society of America*, 63, 1535-1555.

Crichton, R. G. and Fallside, F. (1974). Linear prediction model of speech production with applications to deaf speech training. *Proceedings of Institute of Electrical Engineering Control & Science*, 121, 865-873.

Gersho, A. (1994). Advances in speech and audio processing. *Proceedings of Institute of Electrical and Electronic Engineering*, 82, 900-918.

Guenther, F. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics*, 72, 43-53.

Javkin, H., Baroso, N. A., Das, A., Zerkle, D., Yamada, Y., Murata, N., Levitt, H., and Youdelman, K. (1993). A motivation-sustaining articulatory/acoustic speech training system for profoundly deaf children. *1993 International Conference on Acoustic, Speech and Signal Processing*, 1, 145-148.

Ladefoged, P., Harshman, R., Goldstein, L., and Rice, L. (1978). Generating vocal tract shapes from formant frequencies. *Journal of the Acoustical Society of America*, 64, 1027-1035

- McCandless, S. S. (1974). An algorithm for automatic formant extraction using linear prediction spectra. *Institute of Electrical and Electronic Engineering Transactions on Acoustic Speech and Signal Processing*, 20, 135- 141.
- Mermelstein, P. (1967). Determination of the shape of the human vocal tract from acoustical measurements. *Journal of the Acoustical Society of America*, 41, 1283- 1294.
- Paige, A. and Zue, W. (1970). Computation of vocal tract area functions. *Institute of Electrical and Electronic Engineering Transactions on Audio Electroacoustics*, 18, 7- 18.
- Park, S. H., Kim, D. J., Lee, J. H., and Yoon, T. S. (1994). Integrated speech training system for the hearing impaired. *Institute of Electrical and Electronic Engineering Transactions on Rehabilitation*, 2, 189- 196.
- Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., and Jackson, M. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, 92, 3078- 3096.
- Rahim, M. G., Goodyear, C. C., Kleijin, W. B., Schroeter, J., and Sondhi, M. (1993). On the use of neural network in articulatory speech synthesis. *Journal of the Acoustical Society of America*, 93, 1109- 1121.
- Schroeder, M. R. (1967). Determination of the geometry of the human vocal tract by acoustic measurements. *Journal of the Acoustical Society of America*, 41, 1002- 1010.
- Schroeter, J. and Sondhi, M. M. (1994). Techniques for estimating vocal-tract shapes from the speech signal. *Institute of Electrical and Electronic Engineering Transactions on Speech and Audio Processing*, 2, 133- 150.
- Wakita, W. (1973). Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms. *Institute of Electrical and Electronic Engineering Transactions on Acoustic, speech and signal Processing*, 21, 417- 427.