SEARMA(Simultaneous Estimation of Autoregressive and Moving-average Parameters)法と音声規則合成

森川　博由（福井大学工学研究科）

大学院修士課程時代(1971～1973)

- 学会講演

(1)森川博由, 橋本清, “統計的検定法を用いた有声・無声の弁別法, ”日本音響学会春季研究発表会講演論文集, 3-2-21 (1972-5).

(2)森川博由, 橋本清, “状態空間法による適応形音声分析系, ”日本音響学会秋季研究発表会講演論文集, 2-2-2 (1972-10).

(3)森川博由, “付加雑音に汚染された不規則信号に対する自己成長を持つ多類別機械, ”電子通信学会全国大会論文集, 1311 (1973-3) .

(4)森川博由, “最小二乗法を用いた極及び零点を含む音声分析, ”日本音響学会秋季研究発表会講演論文集, 1-3-14 (1973-10).

(5)森川博由, “音声合成用ディジタルフィルタの分析と合成, ”日本音響学会秋季研究発表会講演論文集, 1-3-15 (1973-10).

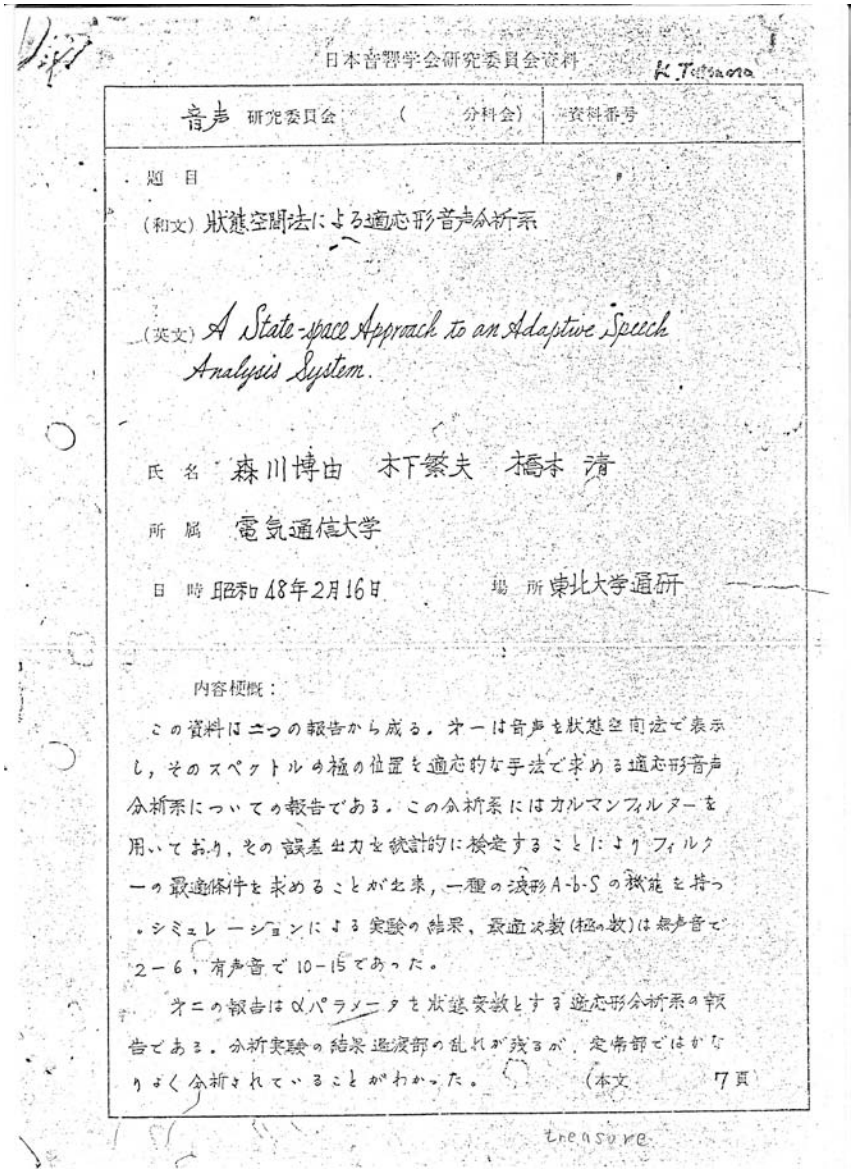
- 研究会資料

(1)森川博由, “カルマンフィルタによる教師なしの適応パターン認識系,”電子通信学会オートマトン・インホメーション理論研究会資料, A71-117, IT-102 (1972-3).

(2)森川博由, 橋本清, “状態空間法による適応形音声分析系, ”日本音響学会音声研究会資料, (1973-2).

音声研究会
日時：昭和48年2月16日
状態空間法に適応形音声分析系
結論：
シミュレーションによる実験の結果、最適次数（極の数）は無声音で2-6、有声音で10-15であった。

参考
板倉文忠、斉藤収三、“統計的手法による音声スペクトル密度とホルマン周波数の推定”,信学論(A)、1970
H.Akaike,“A new look at the statistical model identification ,” lEEE Trans. , vol.AC-19, 1974



日本音響学会研究委員会資料　K. Tohsaka

音声 研究委員会　（　分科会）　資料番号

題目

（和文）状態空間法による適応形音声分析系

（英文）A State-space Approach to an Adaptive Speech Analysis System.

氏名　森川博由　木下繁夫　橋本清

所属　電気通信大学

日時　昭和48年2月16日　　場所　東北大学通研

内容梗概：

この資料は二つの報告から成る。その一は音声を状態空間法で表示し，そのスペクトルの極の位置を適応的な手法で求める適応形音声分析系についての報告である。この分析系にはカルマンフィルターを用いており，その誤差出力を統計的に検定することによりフィルターの最適條件を求めることが出来，一種の波形A-b-Sの機能を持つ。シミュレーションによる実験の結果，最適次数（極の数）は無声音で2－6，有声音で10－15であった。

その二の報告はαパラメータを状態変数とする適応形分析系の報告である。分析実験の結果過渡部の乱れが残るが，定常部ではかなりよく分析されていることがわかった。　（本文　7頁）

treasure

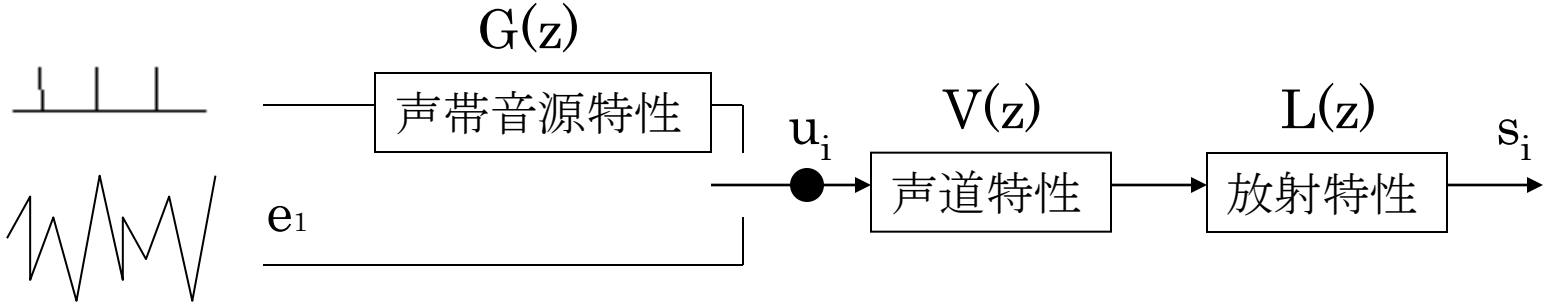


図1　音声生成過程の線形モデル

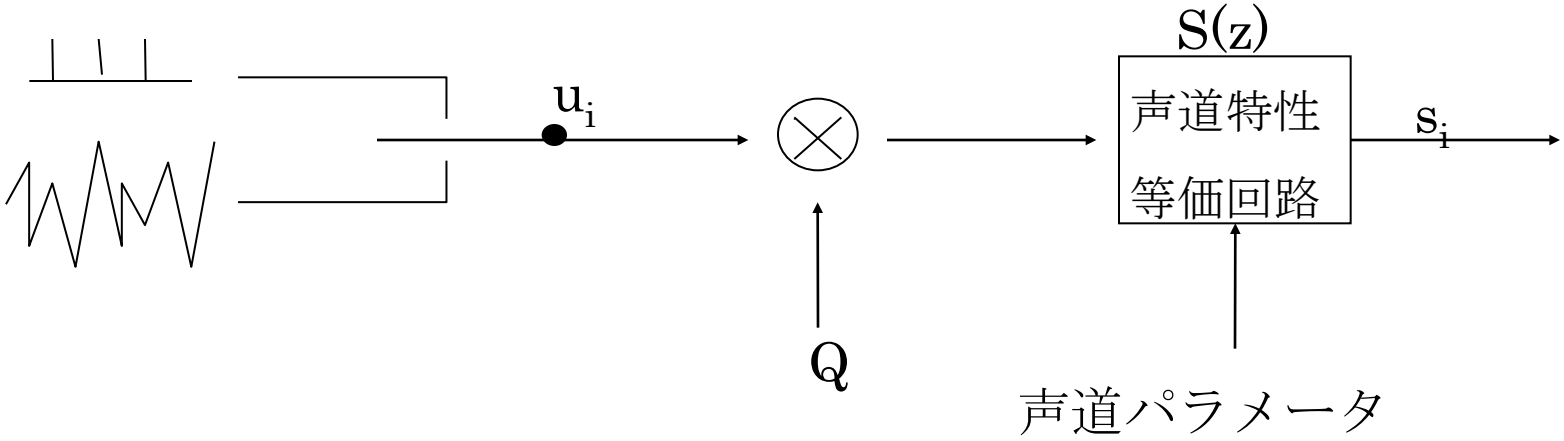


図2　簡略化した音声生成過程の線形モデル

音声分析において解決すべき課題

- 観測区間の影響
  (1)AR･MAパラメータの同時推定法による音声分析
  電子通信学会論文誌，Vol. 61-A, 195-202，1978.3
  (2)SEARMA法による音声分析における観測区間の適応的制御
  日本音響学会誌，39巻，512-520, 1983.8
- モデルの次数の影響
  (1)Adaptive analysis of speech based on a pole-zero representation IEEE Trans. , Vol. ASSP-30, 77-88 , 1982.2
  (2)Adaptive estimation of time-varying model order in the ARMA speech Analysis, IEEE Trans., Vol. ASSP-38, 1073-1083, 1990.7
- 付加雑音の影響
  (1)雑音環境下における音声のスペクトル推定
  電子通信学会論文誌，Vol.J65-A, 987-994, 1982.10
  (2)Spectral estimation of speech corrupted by colored noise
  Traitement du Signal, Vol. 4, 439-445, 1987.10
  (3)Noise reduction of speech signal by adaptive Kalman filtering
  Automatique-Productique Informatique Industrielle, Vol. 22, 53-68, 1988.1

$$s_i + \alpha_1 s_{i-1} + \cdots + \alpha_n s_{i-n} = u_i + \beta_1 u_{i-1} + \cdots + \beta_m u_{i-m}$$

ARMAモデル

SEARMA法の主な特徴

（１）ARMAパラメータの次数とそれらの値を同時に推定することができる 。

（２）事前に観測区間を設定することなしに、パラメータの収束判定により、声道伝達関数の時間的変化に追従した適応的分析を行うことができる。

（３）次数推定に事前確率の概念と損失関数を導入することにより、フレーム相互間での次数の相関を考慮した適応的・連続的次数の推定が可能となる。

評価関数 $I_k(\phi)$

$$I_k(\phi)=\sum_{i=1}^{k}\left\|s_i-S_{i-1}{}^T\phi\right\|+\left\|\phi-\phi_0\right\|^2 P_0^{-1}$$

(最小2乗法によるARMAパラメータの推定法) 評価関数 $I_k(\phi)$ を最小にするφの推定値 $\hat{\phi}$は次の関係式で得られる。

$$\hat{\phi}_{r+1}=\hat{\phi}_k+K_k\left(s_{k+1}-\hat{S}_k{}^T\hat{\phi}_k\right)$$

$$K_k=P_k\hat{S}_k\left(\hat{S}_k{}^T P_k\hat{S}_k+1\right)^{-1}$$

$$P_k=P_{k-1}-P_{k-1}\hat{S}_{k-1}\left(\hat{S}_{k-1}P_{k-1}\hat{S}_{k-1}+1\right)^{-1}\hat{S}_{k-1}{}^T P_{k-1}$$

ここで

$$\hat{S}_{k-1}{}^T=\left[s_{k-n},...,s_{k-1},\hat{u}_{k-m},...,\hat{u}_{k-1}\right]$$

$$\hat{u}_k=\hat{s}_k-\hat{S}_{k-1}{}^T\hat{\phi}_k$$

$$\hat{\phi}^T=\left[-\hat{\alpha}_n,...,-\hat{\alpha}_1,\hat{\beta}_m,...,\hat{\beta}_1\right]$$

である。

（AR・MAパラメータの同時推定法による音声分析, 信学論（A）,
Vol. 61-A, 195-202, 1978.3）

音声生成過程のダイナミックス

$$\phi_k = \Psi \phi_{k-1} + \nu_{k-1}$$

(非定常信号に対するARMAパラメータの推定法)

ϕ_k が式のダイナミックスを持つ時、ϕ_k の最小2乗推定は

$$\hat{\phi}_{k+1} = \Psi \hat{\phi}_k + K_k (s_{k+1} - \hat{S}_k{}^T \Psi \hat{\phi}_k)$$

$$K_k = M_{k+1} \hat{S}_k (\hat{S}_k{}^T M_{k+1} \hat{S}_k + \lambda_k)^{-1}$$

$$M_k = \Psi P_k \Psi^T + Q_k{}^{-1} R_k$$

$$P_{k+1} = \lambda_{k-1}{}^{-1} [M_{k+1} - M_{k+1} \hat{S}_k (\hat{S}_k{}^T M_{k+1} \hat{S}_k + \lambda_k)^{-1} \hat{S}_k{}^T M_{k+1}]$$

$$\hat{u}_k = \hat{u}_k{}^p + \hat{u}_k{}^w$$

で与えられる。

(SEARMA法による音声分析における観測区間の適応的制御
音響学会誌、39巻、512-520、1983.8)

[命題１］ $\hat{\phi}$の出力時点は

$$J_k = \left\| \hat{\phi}_k - \hat{\phi}_{k-1} \right\|^2$$

が n サンプル以上連続して閾値 θ 以下となる時点とする。

[定理１］　命題１のアルゴリズムを適用することにより、$\hat{\phi}$ の推定誤差の共分散行列 $\boldsymbol{\Sigma}$ に対し、観測区間の終端**N**において次の不等式が成立する。

$$\left\| \sum_N \right\| \leq \left(1 - \xi_{\max}^2\right)^{-1} \left\| \Gamma_N \Gamma_N^T \right\| \left\| \mathbf{K}_{N-1} \right\|^{-2} \theta$$

ただし、$\xi_{\max}$ は　$P_{N+1} P_N^{-1}$　の固有値 ξ のうちの最大値、Γ_N は　$P_{N+1} \hat{S}_N$ である。

（SEARMA法による音声分析における観測区間の適応的制御音響学会誌, 39巻，512-520, 1983,8）

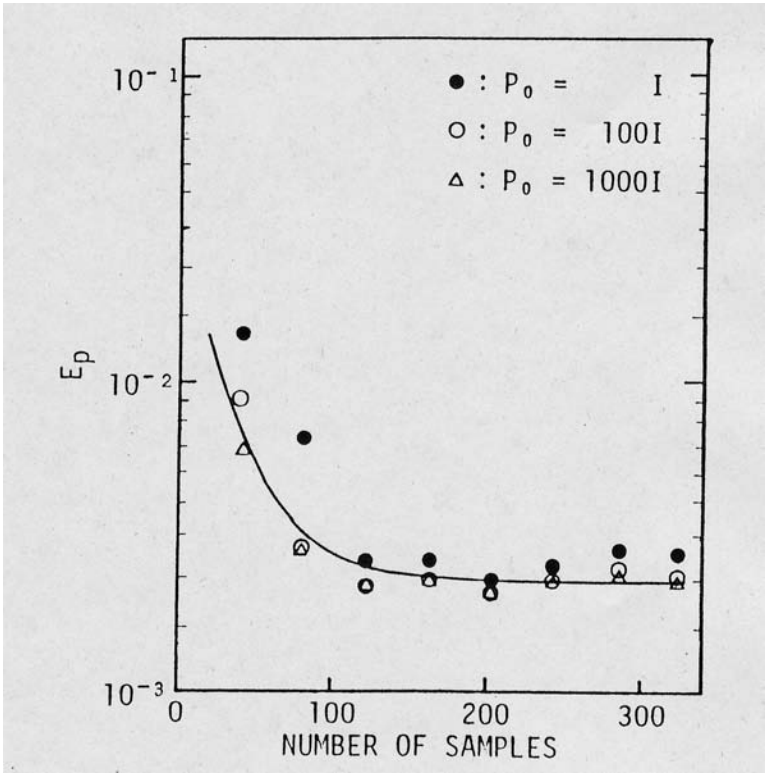


Fig.3. Effect of the variance of P_0 on the convergence of estimated pole frequencies for synthetic vowel /i/ with impulse excitation. I denotes unit matrix.

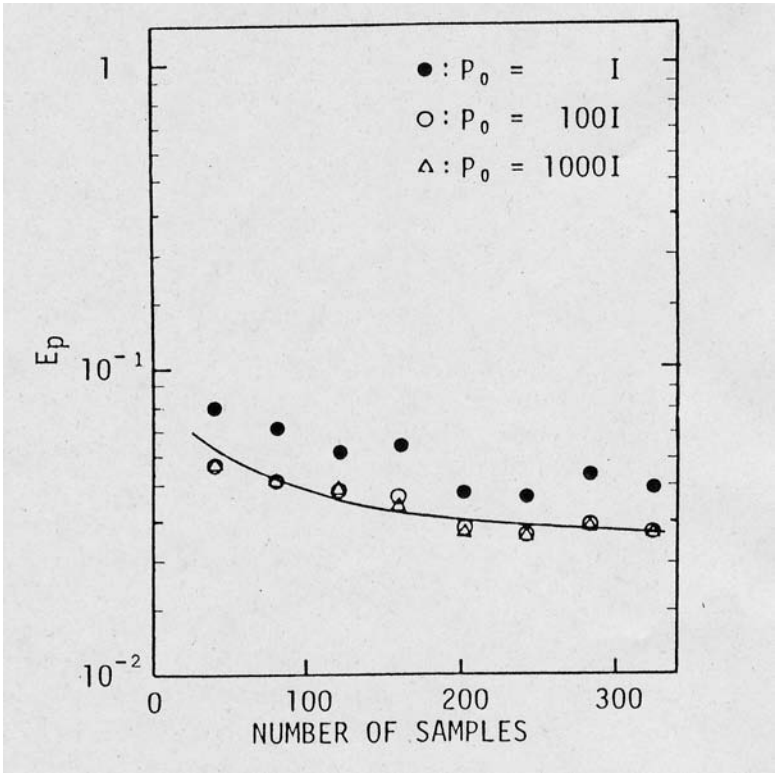


Fig.4. Effect of the variance of P_0 on the convergence of estimated pole frequencies for synthetic vowel /i/ with noise excitation. I denotes unit matrix.

(Adaptive analysis of speech based on a pole-zero representation IEEE Trans. , Vol. ASSP-30, 77-88 , 1982.2)

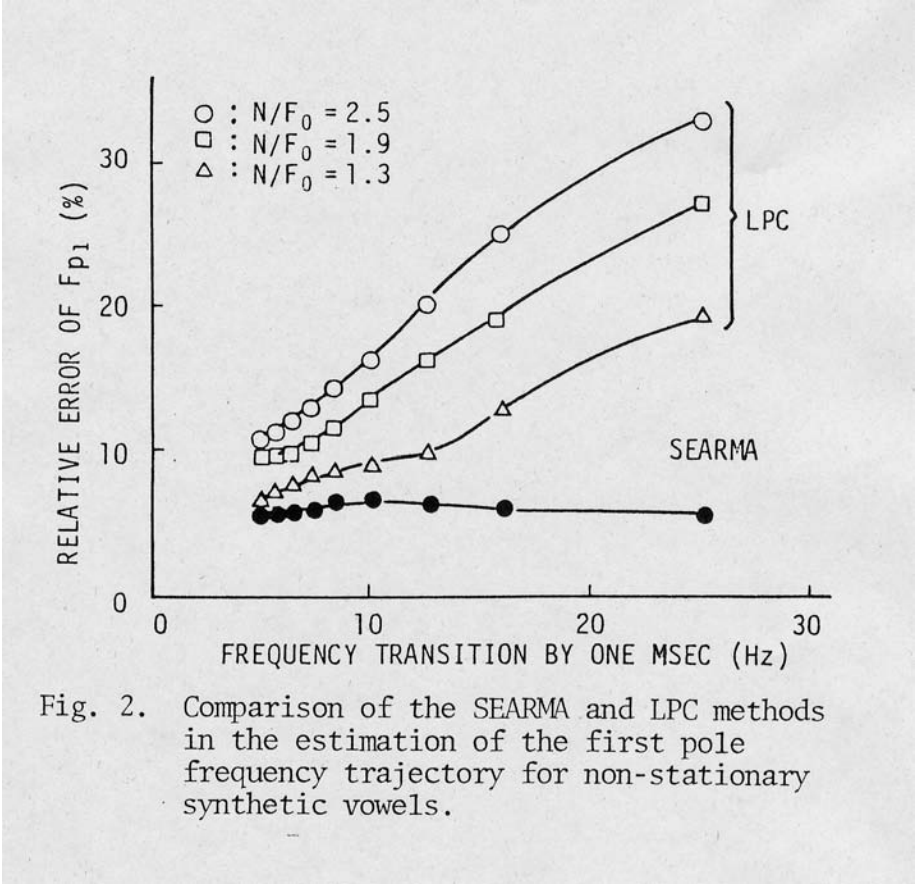


Fig. 2. Comparison of the SEARMA and LPC methods in the estimation of the first pole frequency trajectory for non-stationary synthetic vowels.

（SEARMA法による音声分析における観測区間の適応的制御、
音響学会誌, 39巻, 512-520, 1983,8）

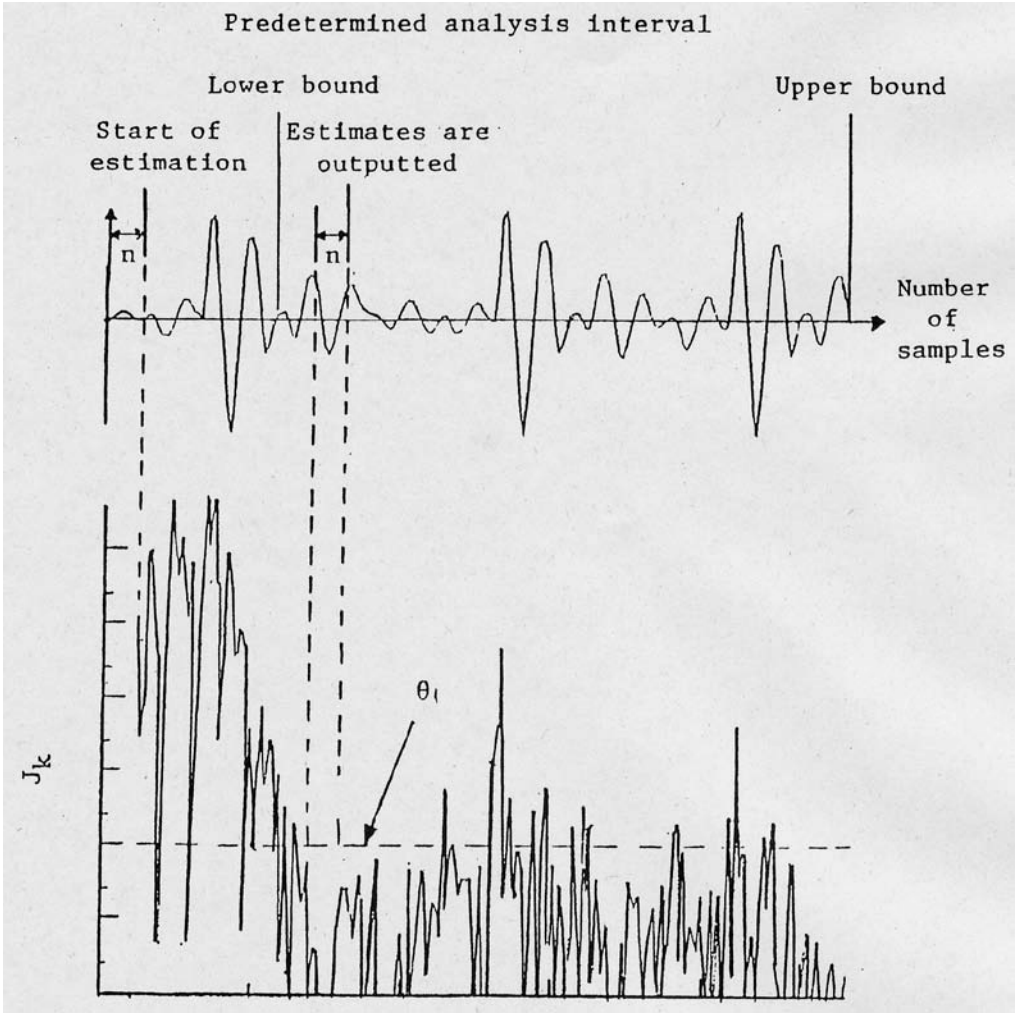


Fig.2. Schematic illustration for determination of end of analysis interval in the SEARMA algorithm. θ denotes the threshold for convergence of estimation.

(Adaptive analysis of speech based on a pole-zero representation IEEE Trans. , Vol. ASSP-30, 77-88 , 1982.2)

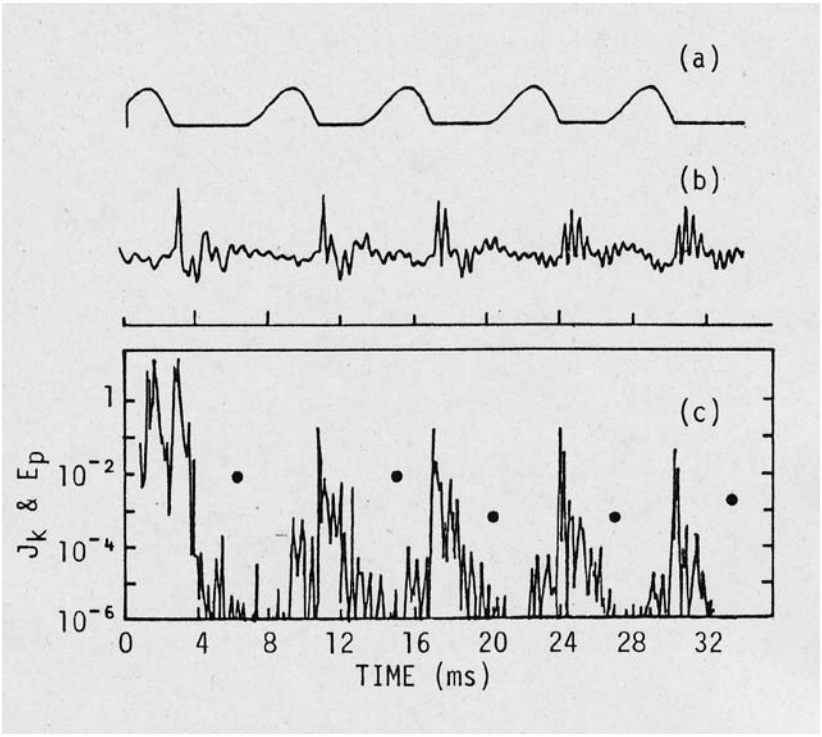


Fig.14. Illustration of (a) simulated glottal waveform, (b) waveform of transition from /a/ to /i/, and (c) J_k and E_p(●) obtained by the generalized SEARMA method.

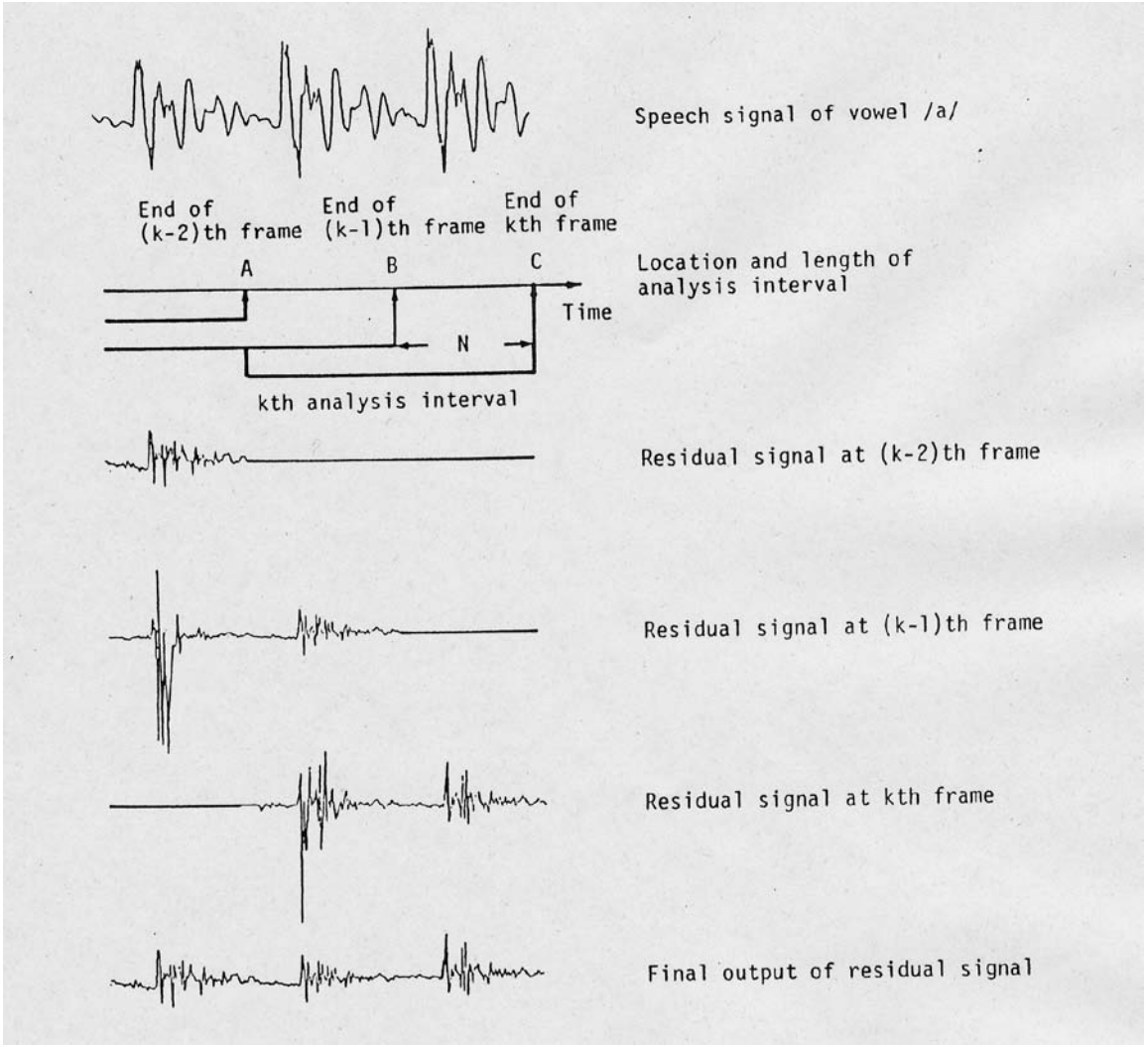


Fig.2. Illustration of the location and the length of the analysis interval.

（Adaptive estimation of time-varying model order in the ARMA speech Analysis, IEEE Trans., Vol. ASSP-38, 1073-1083, 1990.7)

（森川、１９７２、１９７６）

$$SSR(n,m) = 10\log_{10}\frac{b_N}{r_N}$$

（Akaike、１９７４）

$$AIC(n,m) = N\log r_N + 2(n+m)$$

（Rissanen、１９７８）

$$MDL(n,m) = N\log r_N + (n+m)\log N$$

（Hannan and Quinn、１９７９）

$$HQ(n,m) = N\log r_N + (n+m)c\log\log N$$

（Dukkila and Krishnaiah、１９８８）

$$AM(n,m) = N\log r_N + (n+m)g(N)$$

（Adaptive estimation of time-varying model order in the ARMA speech Analysis, IEEE Trans., Vol. ASSP-38, 1073-1083, 1990.7）

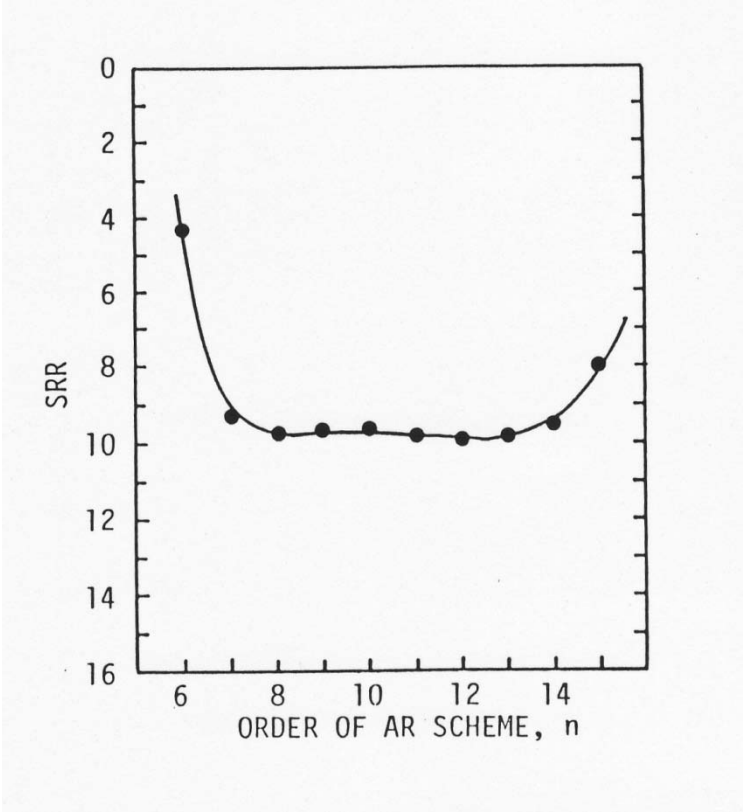


Fig.11. SRR versus order of AR model for synthetic nasal consonant /m/ by an ARMA model (n=8,m=4) with impulse excitation.

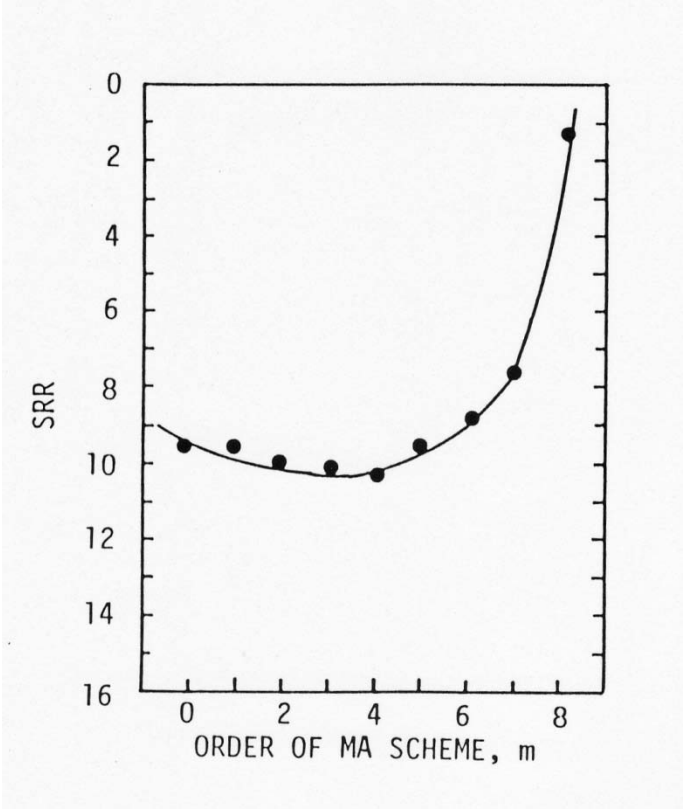


Fig.12. SRR versus of MA scheme of ARMA model (n=8) for synthetic nasal consonant /m/by ARMA model (n=8,m=4) with impulse excitation.

(Adaptive analysis of speech based on a pole-zero representation
IEEE Trans. , Vol. ASSP-30, 77-88 , 1982.2)

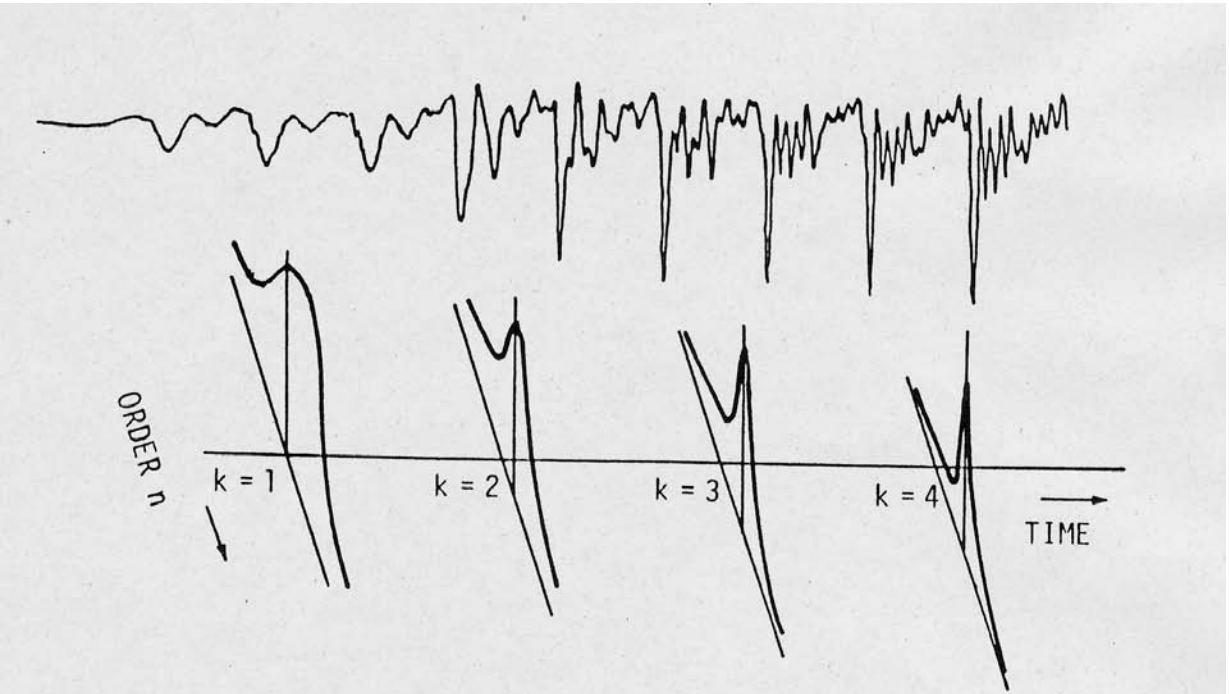


Fig. 1. Transitions of conditional probability density functions $p(n,m|S_k^N)$ (scalar case for only n) for multistage order estimation.

(Adaptive estimation of time-varying model order in the ARMA speech Analysis, IEEE Trans., Vol. ASSP-38, 1073-1083, 1990.7)

（統計的決定論に基づく次数推定法） $S_k{}^N$ の次数(n,m)は $D(n,m)$で与えられた平均損失が最小となるように決定される。すなわち、

$$\min_{i,j} D(i,j) = d(n,m)$$

である。

$$D(n,m) = \sum_{i=\underline{n}}^{\bar{n}} \sum_{j=\underline{m}}^{\bar{m}} c(i,j \mid n,m)\, p(i,j \mid S_k{}^N)$$

the model conditional likehood ratio $\Lambda(n,m \mid S_k{}^N)$

$$\Lambda(n,m \mid S_k{}^N) = \frac{f(S_k{}^N \mid n,m)}{f(S_k{}^N \mid 0,0)}$$

$$p(n,m \mid S_k{}^N) = \frac{\Lambda(n,m \mid S_k{}^N)\, p(n,m)}{\sum_{i=\underline{n}}^{\bar{n}} \sum_{j=\underline{m}}^{\bar{m}} \Lambda(i,j \mid S_k{}^N)\, p(i,j)}$$

$$\Lambda(n,m \mid S_k{}^N) = \frac{\prod_{i-1}^{N} \frac{1}{(2\pi)^{1/2} r_i^{1/2}} \exp\left(-\frac{1}{2} w_i{}^2 r_i^{-1}\right)}{\prod_{i-1}^{N} \frac{1}{(2\pi)^{1/2} b_i^{1/2}} \exp\left(-\frac{1}{2} s_i{}^2 b_i^{-1}\right)}$$

(Adaptive estimation of time-varying model order in the ARMA speech Analysis
IEEE Trans., Vol. ASSP-38, 1073-1083, 1990.7)

損失関数

$$C(n,m \mid n,m) = 0$$

$$C(i,j \mid n,m) = 1 + C_1 + C_2 + C_3$$

$$(i,j) \neq (n,m)$$

ただし

$$C_1 = G_1\left(\frac{i}{n} + \frac{j}{m}\right)$$

$$C_2 = G_2\left(\frac{|i - \hat{n}_{k-1}|}{n} + \frac{|j - \hat{m}_{k-1}|}{m}\right)$$

$$C_3 = G_3\,\frac{(\overline{m} - j)}{m}$$

（Adaptive estimation of time-varying model order in the ARMA speech Analysis, IEEE Trans., Vol. ASSP-38, 1073-1083, 1990.7）

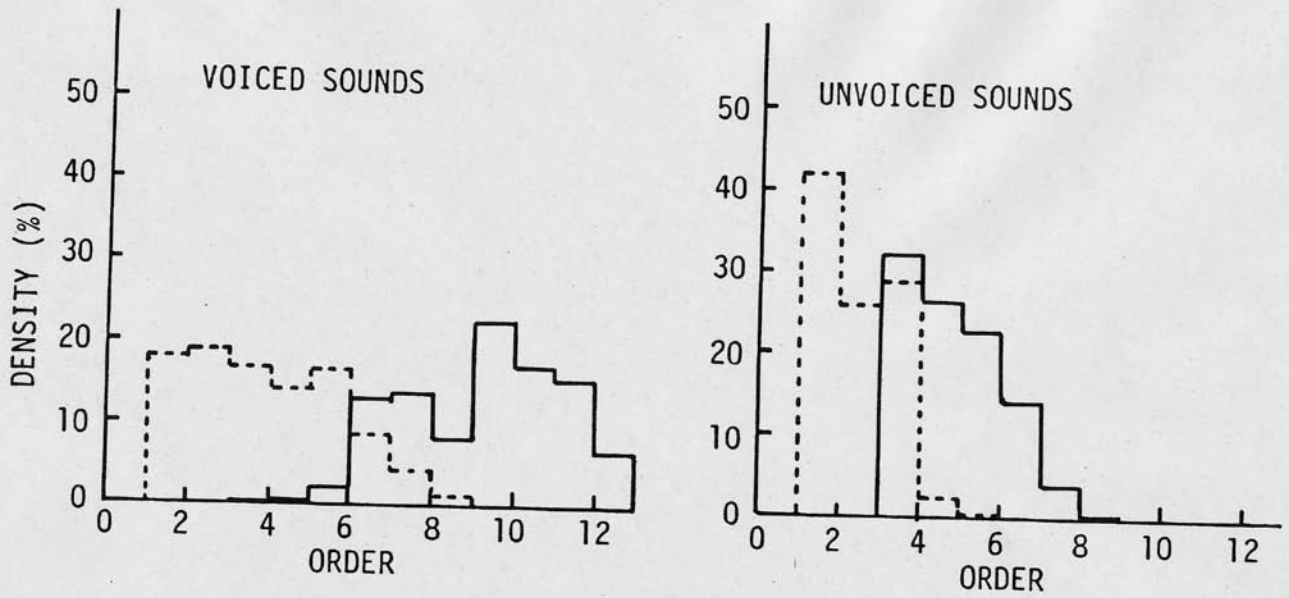


Fig. Density distributions of orders n (straight line) and m (dotted line) of ARMA model for Japanese voiced and unvoiced sounds.

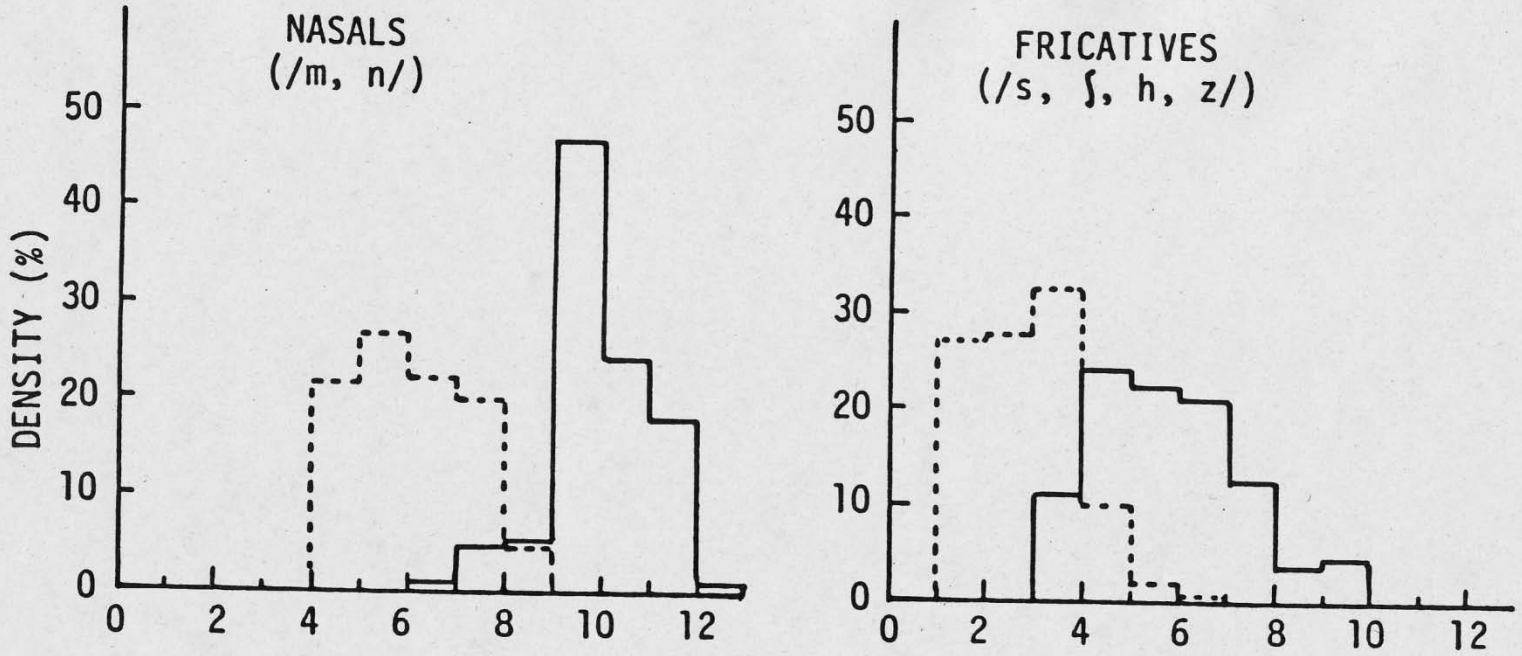


Fig. Density distributions of orders n (straight line) and m (dotted line) of ARMA model for Japanese nasals and fricatives.

(Adaptive estimation of time-varying model order in the ARMA speech Analysis, IEEE Trans., Vol. ASSP-38, 1073-1083, 1990.7)

音声発達過程を模擬する音声規則合成のための要素技術

- 分析方法(ARMAモデルに基づく音声分析)
  - SEARMA法
- 合成方法(フォルマント・アンチフォルマント合成)
  - ARMA合成器(極・零パラメータ→ARMAパラメータ)
- 音韻パラメータ、韻律パラメータのモデル化
  - 平滑化スプライン関数によるパラメータ変化パターンのモデル化
- 音声言語発達過程の規則化
  - 絵カード、文字カードによる調音テストによる音声データの蓄積
  - 調音発達過程の規則化
  - アクセント発達過程の規則化
- データベース化
  - 各年齢における音声波形のデータベース
  - 各年齢における音素・音節・単語の各パラメータ変化パターンの節点(パラメータの値、時間)のデータベース

持続時間の分析

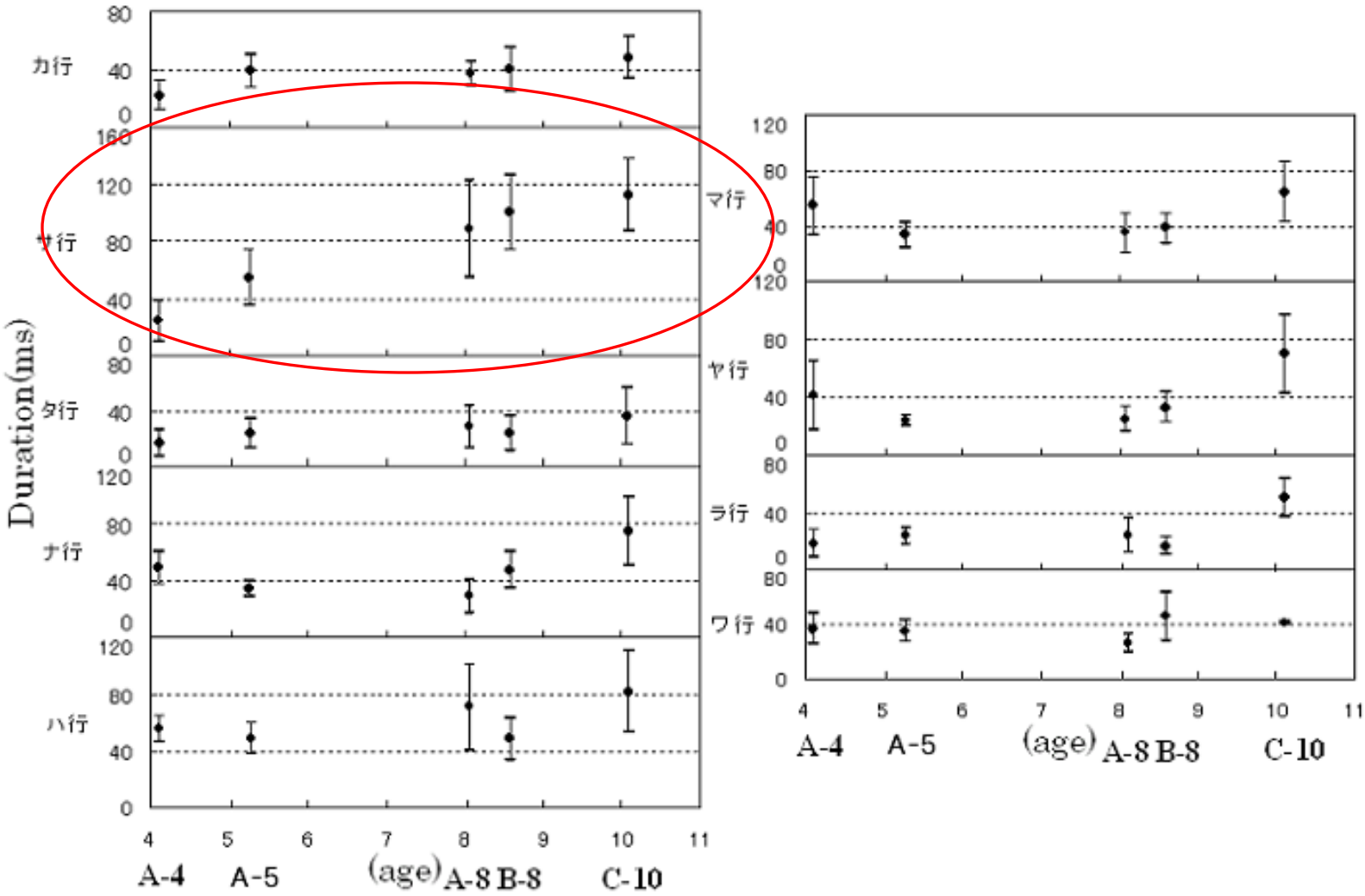


図1．各テスト語における語頭子音の持続時間の分布と経年変化

アクセントの分析

表2．アクセント型ごとの経年変化

	4歳		5歳		8歳	
	ピッチパターン	知覚実験	ピッチパターン	知覚実験	ピッチパターン	知覚実験
平板型(21語)	17語(81%)	13語(62%)	18語(86%)	17語(81%)	20語(95%)	19語(90%)
尾高型(5語)	4語(80%)	3語(60%)	3語(60%)	4語(80%)	5語(100%)	4語(80%)
中高型(5語)	3語(60%)	1語(20%)	3語(60%)	3語(60%)	4語(80%)	3語(60%)
頭高型(18語)	9語(50%)	8語(44%)	14語(78%)	8語(44%)	15語(83%)	11語(61%)

規則化

・持続時間について
  ・サ行以外の子音（分析結果を使用）
    それぞれの年齢ごとの持続時間を使用
  ・サ行の持続時間（規則化）
    年齢ごとに持続時間を設定し制御

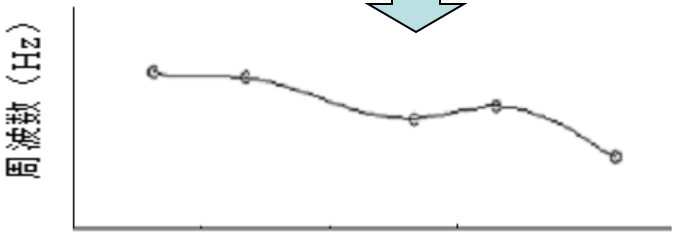
・アクセントについて
  ・振幅強度をスプラインでモデル化（規則化）
  ・ピッチパターンをスプラインでモデル化（規則化）

・フォルマントについて
  ・年齢ごとの各データからパラメータを取り出す
  （分析結果を使用）
  ・スプラインでモデル化（規則化）

規則合成の過程

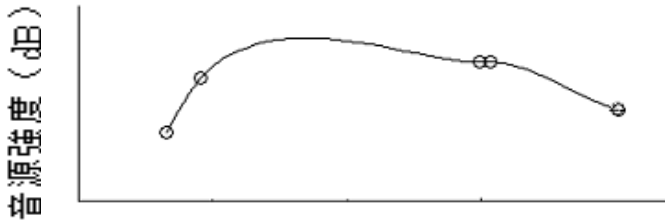
韻律パラメータ
データベース

音韻パラメータ
データベース



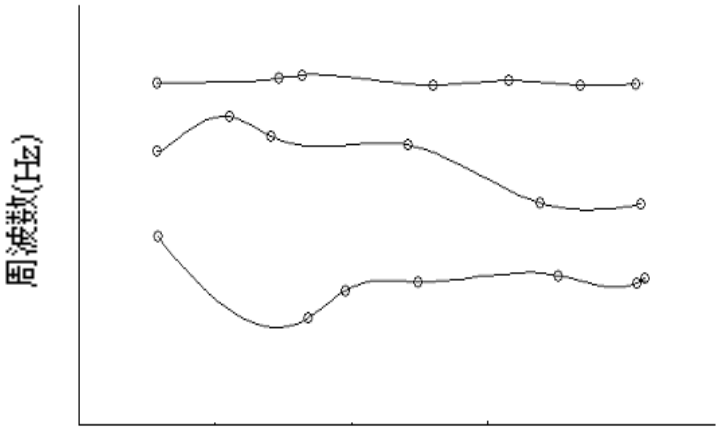
時間
（ms）

基本周波数変化パターン



時間
（ms）

音源強度変化パターン



時間
（ms）

フォルマント周波数変化パターン

図2．データベースからのスプライン関数の再現

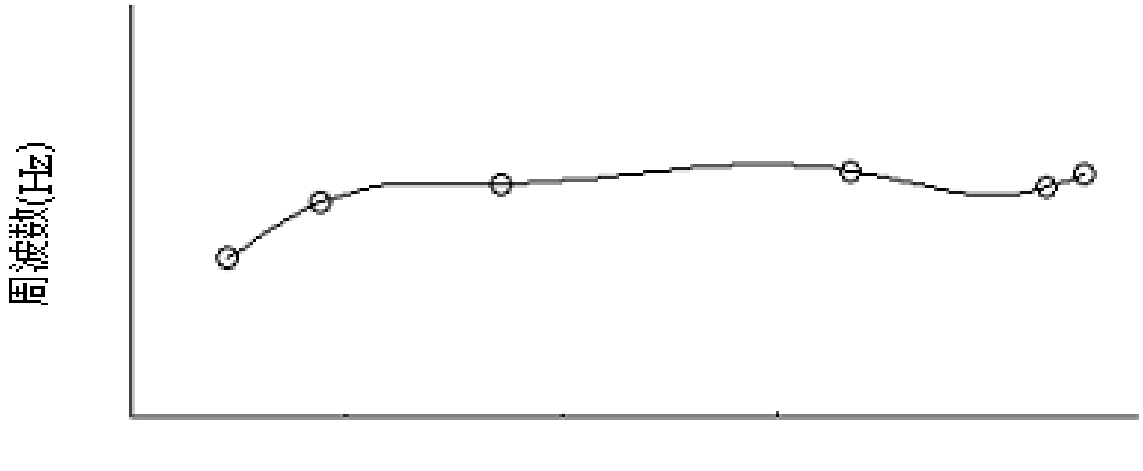


図7．接点をもとに作成した第1フォルマント周波数変化パターン

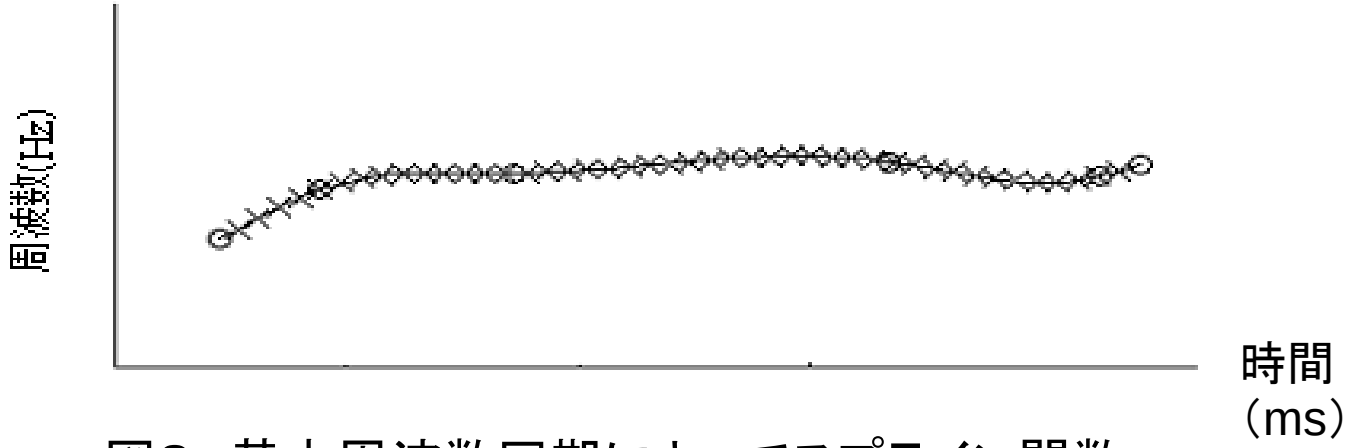


図8．基本周波数同期によってスプライン関数上に得られた点

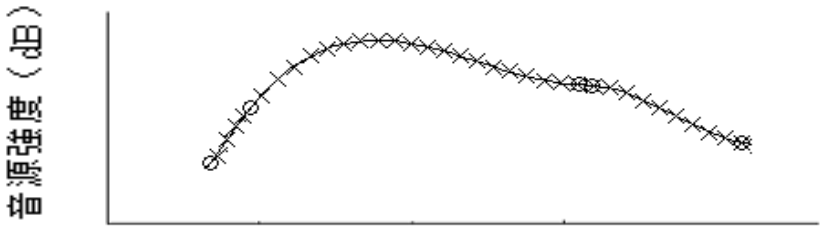


図3．基本周波数同期による、有声音声に対する変化パターンの生成

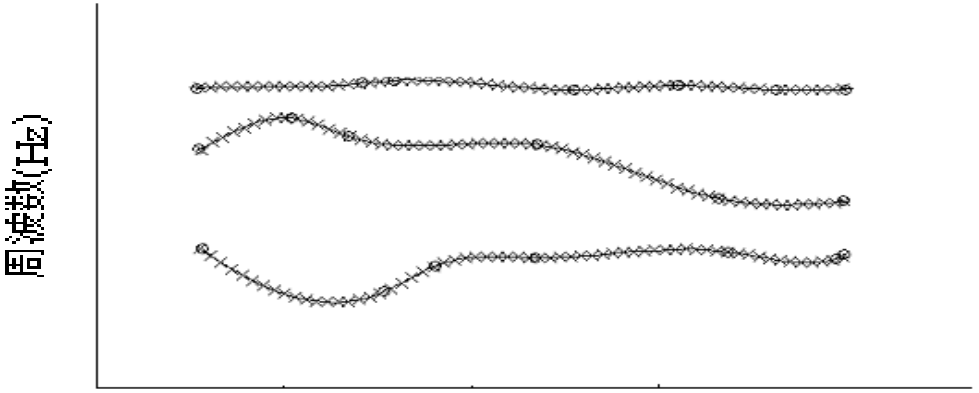
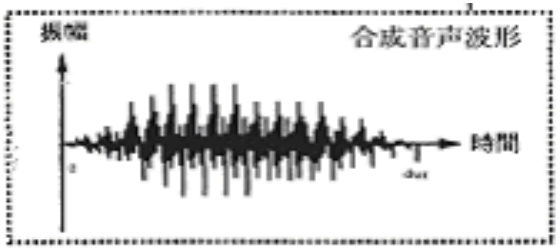


図4．基本周波数同期による、フォルマント周波数軌跡の生成

ARMA合成器



デモンストレーションの内容

- データベースから4歳、5歳、6歳、8歳時の「サル」の各パラメータの節点情報の検索
- 音源信号の生成
  - 振幅変化パターンの生成
  - ピッチ変化パターンの生成
- ARMA合成器パラメータの生成
  - フォルマント周波数変化パターンの生成
  - フォルマント帯域幅変化パターンの生成
- 合成音声生成
  - 合成音声の自然性
  - 音声発達過程の変化を模擬

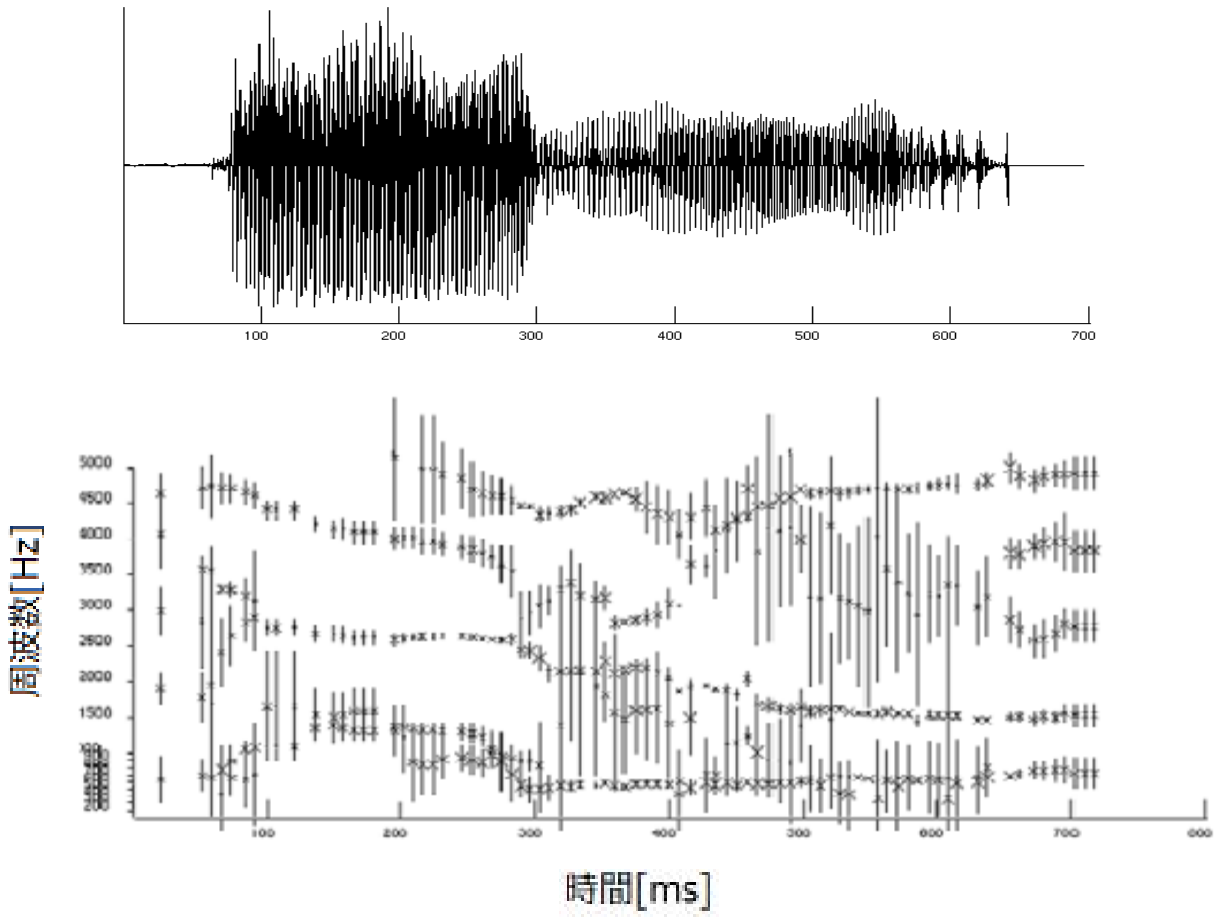


図　次数11のARモデルによる「サル」の4歳時のフォルマント軌跡

合成音声の精度実験

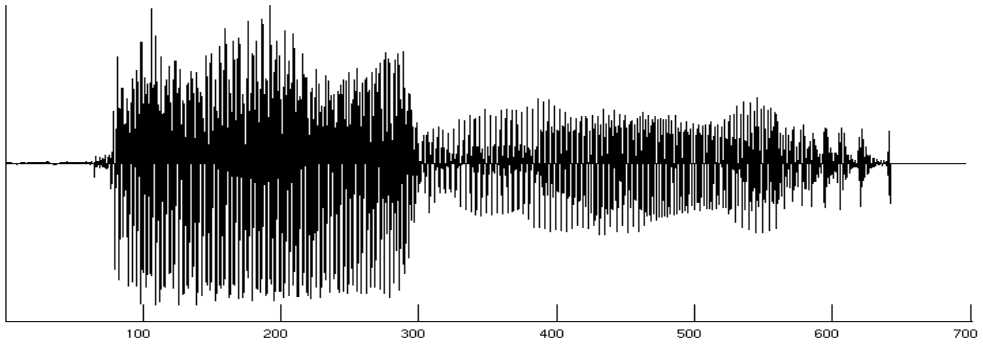


図9．４歳女児の自然音声/saru/の波形

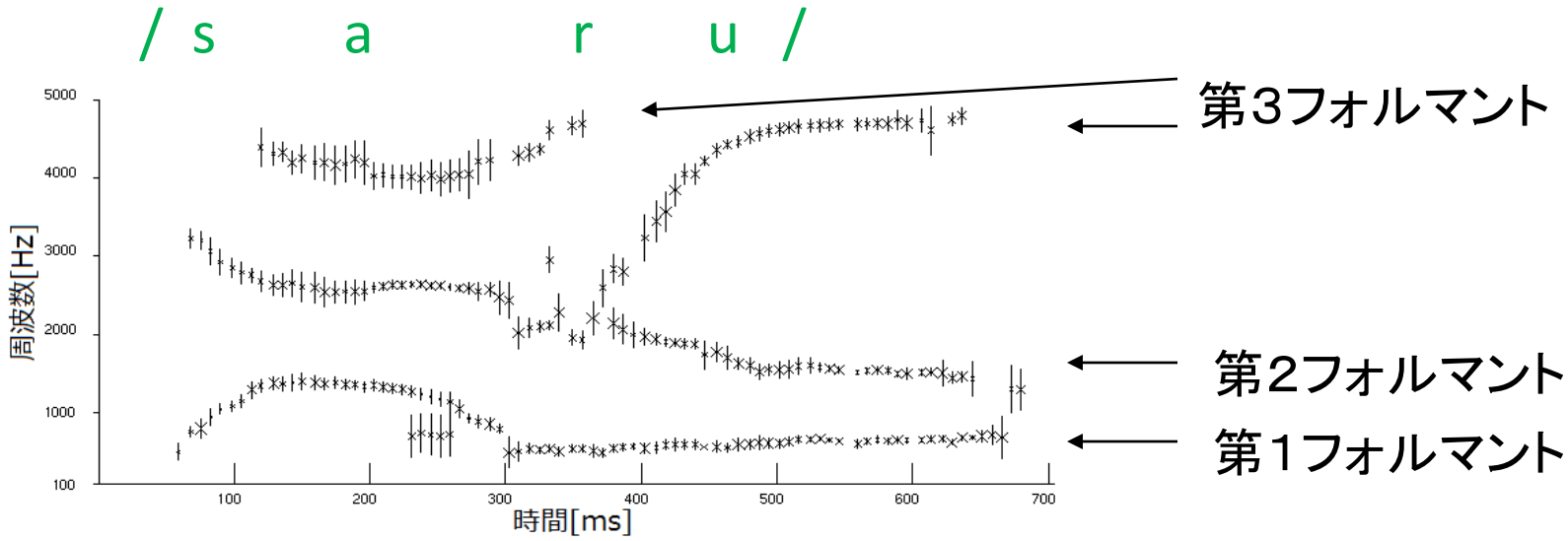


図10　４歳女児の自然音声/saru/のフォルマント周波数軌跡

/ s　　a　　r　　u /

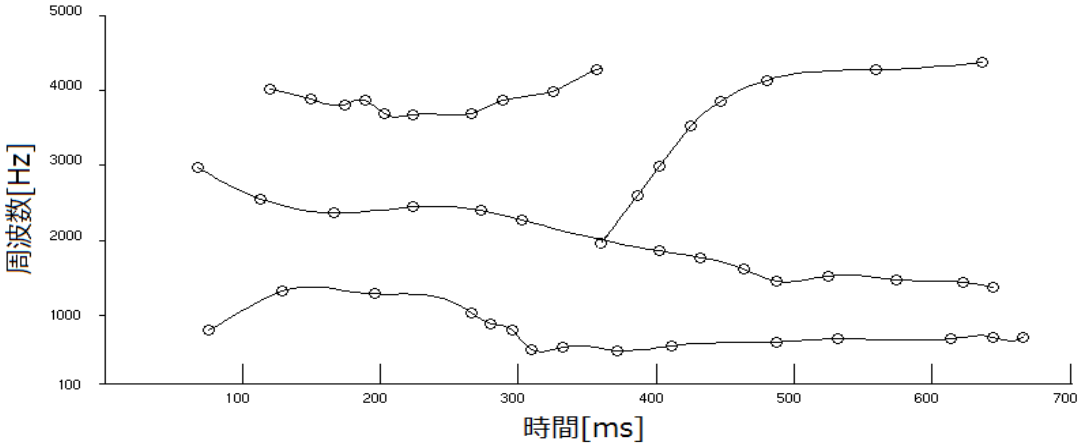


図11　DBより取り出した節点を用いて再現した
４歳女児の/saru/のフォルマント周波数軌跡

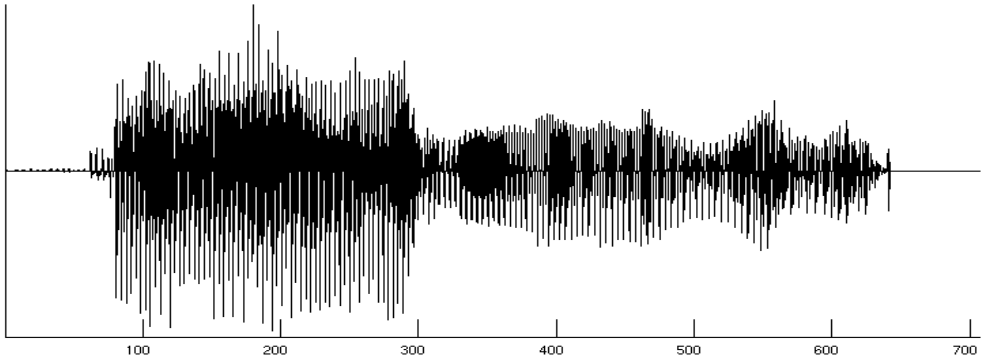


図12．４歳女児の合成音声/saru/の波形

合成音声の発達的変化

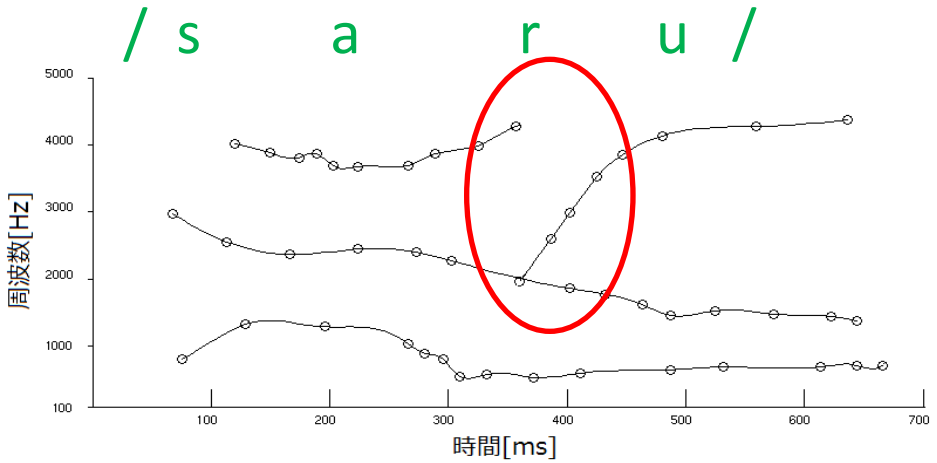


図13　DBより取り出した節点を用いて再現した
4歳女児の/saru/のフォルマント周波数軌跡

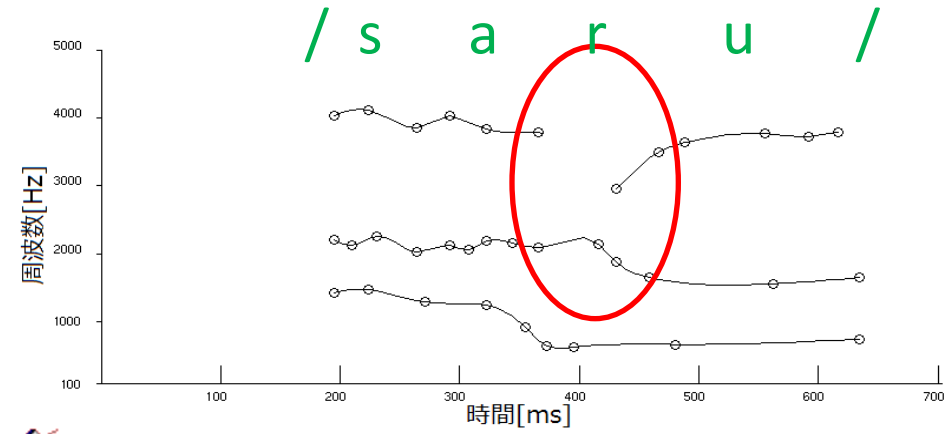


図14　DBより取り出した節点を用いて再現した
5歳女児の/saru/のフォルマント周波数軌跡

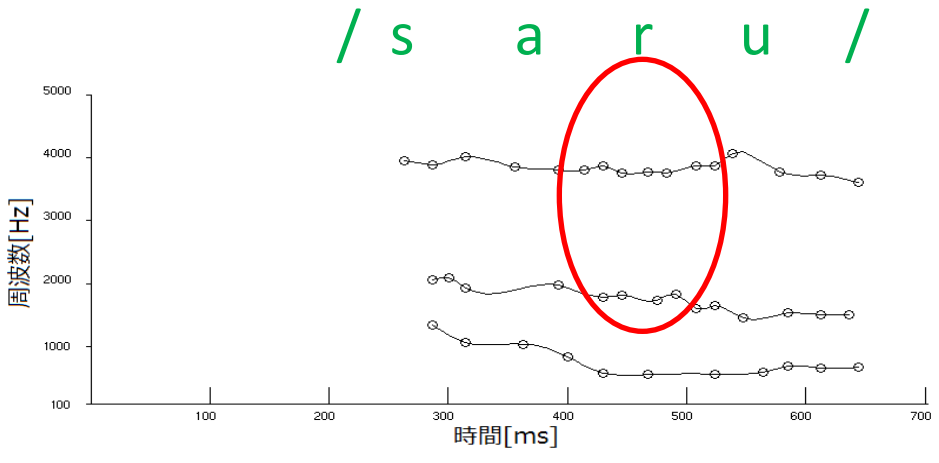


図15　DBより取り出した節点を用いて再現した
6歳女児の/saru/のフォルマント周波数軌跡

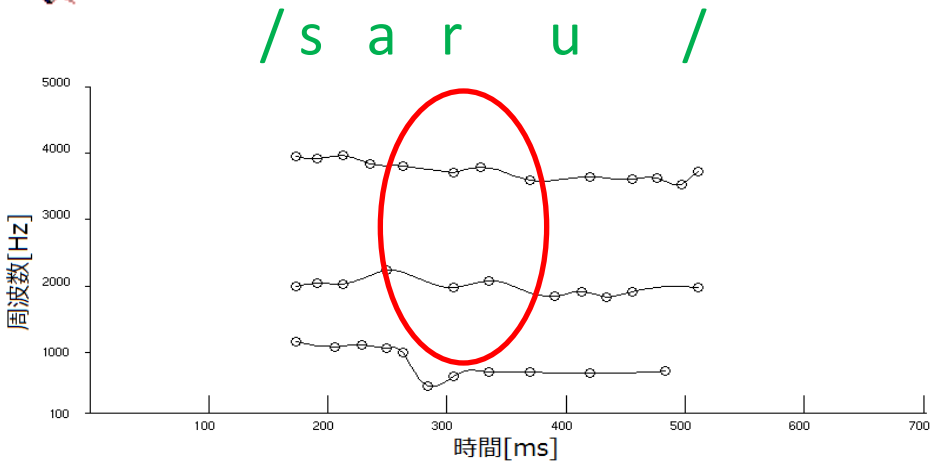


図16　DBより取り出した節点を用いて再現した
8歳女児の/saru/のフォルマント周波数軌跡

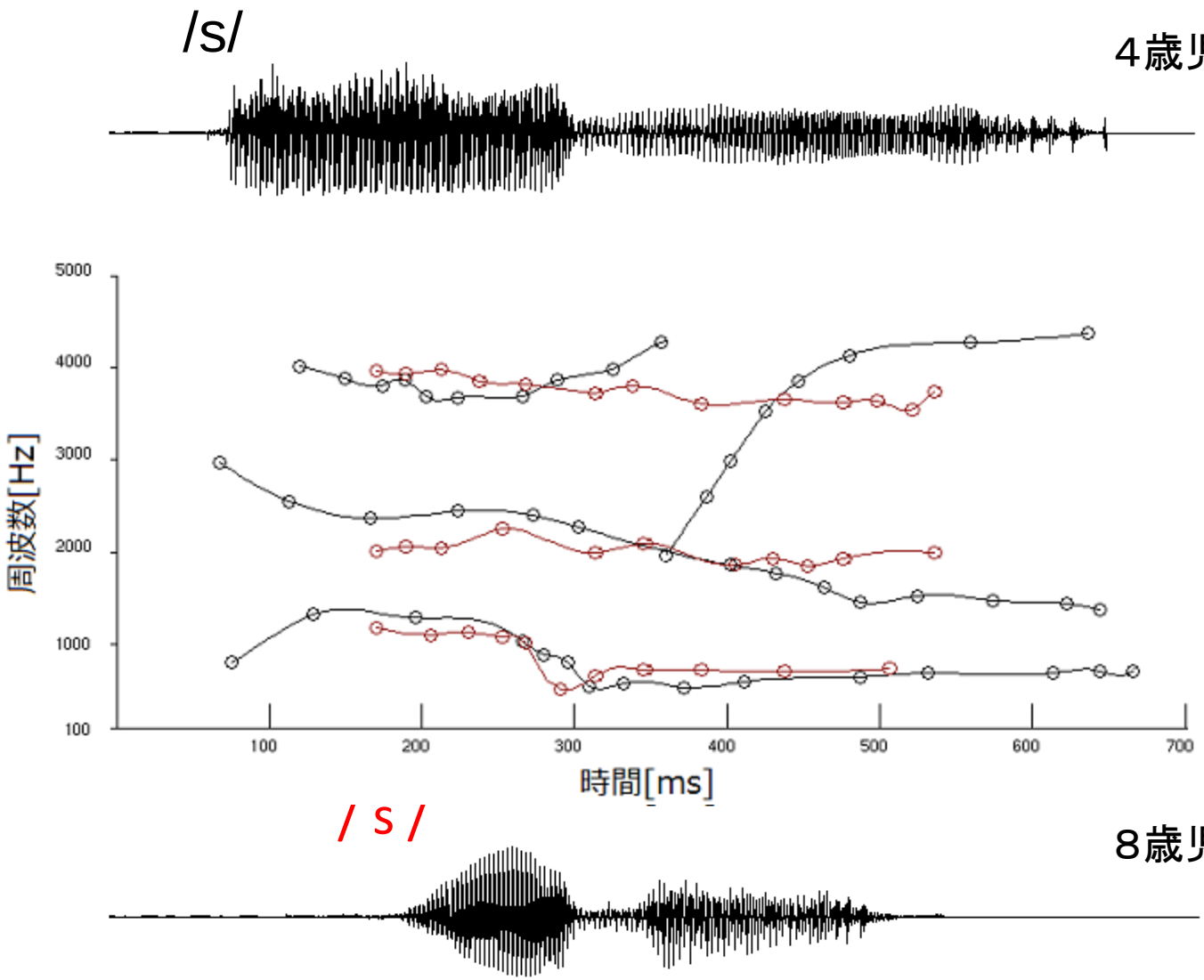


図　DBより取り出した節点を用いて再現した
4歳時（黒）と8歳時（赤）の/saru/のフォルマント周波数軌跡の比較

今後の課題と展望

- データベースの整備
  - 音声波形
  - 音素・音節・単語の各パラメータ
- 音声生成と音声知覚の関係解明
  - 韻律パラメータ変化パターンと音声知覚の関係
  - 音韻パラメータ（フォルマント・アンチフォルマント）変化パターンと音声知覚の関係
- 応用
  - テキストから音声発達過程を模擬する音声の生成
  - 歌声の生成
  - 調音障害者の調音訓練
  - など

- ４自然
- 4合成
- ５合成
- ６合成
- ８合成

糖尿病地域連携医療のミニマムデータセットとデータベースネットワークに関する公開セミナー

主催：日本糖尿病情報学会

日時：2011年4月30(土）13:00～17:00

会場：福井大学アカデミーホール

地域医療連携の課題と展望

●ITによる地域医療連携の課題と展望　　田中博

●富山県における地域医療連携の課題と展望（仮題）　戸邊一之

●石川県における地域医療連携の課題と展望（仮題）　小泉順二

●福井県における地域医療連携の課題と展望（仮題）　番度行弘

日本糖尿病情報学会からの提言

●糖尿病情報連携のミニマムデータセットに関する公開シンポジウムの意見集約と提言　平井愛山

●糖尿病地域連携医療におけるデータベースネットワーク構想　　森川博由

●Net-SMBGシステムと健康増進PHRシステムについて　山崎貞人

総合討論（ミニマムデータセット集約）