ARTICLE TITLE:
Discrimination and clustering for fundamental frequency patterns of infant and parent speech based on time series regression models

AUTHORS:
Hiroko Kato, Ph.D. ♣
Masanobu Taniguchi, Dr. *
Tomohiro Nakatani, Ph.D. ♣
Shigeaki Amano, Ph.D. ♣
♣NTT Communication Science Laboratories, NTT Corporation
*Waseda University

AUTHORS’ FOOTNOTE:
Hiroko Kato is Research Scientist, NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237 Japan (e-mail: katohi@cslab.kecl.ntt.co.jp): Masanobu Taniguchi is Professor, Department of Mathematical Science and Engineering, Waseda University, Tokyo, 169-8555, Japan (e-mail: taniguchi@waseda.jp); Tomohiro Nakatani is Research Scientist, NTT Communication Science Laboratories, NTT Corporation (e-mail: nak@cslab.kecl.ntt.co.jp); Shigeaki Amano is Senior Research Scientist, NTT Communication Science Laboratories, NTT Corporation. (e-mail: amano@cslab.kecl.ntt.co.jp). The first author is grateful to Dr. Parham Zolfaghari of NTT Communication Science Laboratories for helpful advises to numerical representation and Matlab® programming.
ABSTRACT

Fundamental frequency (F0) patterns reflect developmental changes of speech produced by infant and parents. Previous studies on infants’ and parents’ speech showed that the F0 patterns can be classified into patterns such as rising-, flat-, bell-shaped- and complex-shapes. However, these categories were subjectively specified and each F0 pattern was manually classified. Objective and systematic statistical methods have never been applied to F0 pattern classification. For F0 pattern analysis, we propose a statistical method that involves the following steps: 1) introducing a class of time series regression models for F0 data, which includes the missing F0 observations that correspond to the silent parts of speech; 2) fitting the model to the data and estimating the trend based on the regression model; 3) introducing a two-category discrimination method to the estimated F0 trend and classifying the F0 pattern of infants’ and parents’ speech; and 4) comparing the empirical distributional properties of the discrimination function between the infant and the parents. For the proposed method, we proved that misclassification probabilities converge to zero as the length of data tends to infinity, and derived the asymptotic distribution of the discrimination function. We applied the method for analyzing conversations between an infant and parents in a natural environment, spanning the first five years of the infant’s life. The various F0 patterns were automatically classified into such pattern categories as flat, rising, falling, and bell-shaped, and these results demonstrate that the way the infants processed language information changed significantly at around two years of age.

Keywords: Time series regression model with missing data, AIC, F0 patterns, Classification, Discrimination for time series, Language acquisition

1. INTRODUCTION

In the mechanism of human speech, the vocal tract, the region from the vocal cords to the lips, modulates the resonant characteristics of speech timbre. The vibration frequency of the vocal cords is called the fundamental frequency (F0), which determines pitch. Fundamental frequency fluctuation depends on gender and age, and it also includes prosodic information that dictates rules of accent, intonation, rhythm, emotion or intention. Since infant F0 fluctuation resembles human speech, which develops in parallel to human growth, F0 characteristics reflect the sound acquisition process. Infants in particular, whose linguistic ability and vocal organs are underdeveloped, acquire complex F0 patterns in conjunction with the acquisition of vocabulary and sentence structure. Hence, F0 changes in infant speech certainly reflect their development and growth.

Many studies based on speech analysis in the field of developmental psychology have investigated the relationship between the changes in infant speech and the growth of the infant. For example, the typical studies of Fernald and Simon (1984), Papoušek and Papoušek (1989), Siegel et al. (1990), and Davis et al. (2000) are representative of F0 analysis-related research to confirm whether an infant generally imitates the parents’ speech. Their analysis can be summarized as follows:

1. It was based on a small number of sound parameters.
2. The F0 patterns were manually categorized into pre-decided patterns (rising, falling, bell-shaped, flat, complex) by arbitrary established criteria.
3. Correlation analysis was based on a combination of typical sound parameters.

In contrast to previous research, we wish to focus on an objective and systematic statistical analysis of F0 patterns and their classification, and explain the similarities in changes between infants and parents.

In the areas of data mining and rule discovery, there are many clustering methods that can be broadly classified into either categories of whole clustering or subsequence clustering, as reviewed in Keogh et al. (2003). They treated an enormous amount of sequential data as time series data, even when all of it had been divided into subsequent series. However, speech data usually consists of voiced and unvoiced parts. When a subject utters a sentence, the F0 pattern of the unvoiced part is randomly represented as a missing observation (Figure 1). Also, because the F0 data appear to be non-stationary, applying a J-divergence measurement (Kullback, 1987) based on the spectral density of the data, which assumes a stationarity for the data, is insufficient.

In the field of time series analysis, there has been much discussion on discriminant analysis for time series. Most of the discriminant analyses were based on disparity measures between the spectral densities for stationary processes (e.g., Shumway and Unger (1974), Shumway (1982), Zhang and Taniguchi (1994, 1995), Kakizawa et al. (1998), Hallin et al. (1999), Shumway and Stoffer (2000), Taniguchi and Kakizawa (2000)). For locally stationary processes, Sakiyama and Taniguchi (2004) and Huang et al. (2004) discussed discrimination problems by use of disparity measures for time-varying spectra. In this paper, because we observe non-stationary data with missing points, we propose a class of regression time series models with missing observations. We then estimate the trend of the F0 pattern for the entire duration of the time series by a least-squares method, and classify the pattern similarities for each utterance by applying a two-category discriminant analysis to the F0 trends. The similarities among the discrimination function scores directly reflect the pattern similarity. From a theoretical point of view, we prove that misclassification probabilities converge to zero if the length of time series tends to infinity, and derived the asymptotic distribution of the discrimination function. We also conduct a simulation study on our proposed statistical procedure to demonstrate its performance. In the actual data analysis, we employ a database of recorded conversations between an infant and parents, spanning the first five years of the infant’s life. The histogram of the discrimination function values describes the changes in the characteristics of the F0 pattern between infant and parents. Hence, the empirical distributional properties of the discrimination function are provided. Through the statistical methods, we can obtain reasonable results from the perspective of the developmental psychology, that is, the F0 pattern similarities between infant and parents certainly drastically change at around two years of age. These results demonstrate that the way in which the infants processed language information changed significantly at around two years of age.

The remainder of this paper is organized as follows. In Section 2, we explain our infant speech database. In Section 3, we introduce our proposed statistical procedure that combines a regression time series model with missing observations and a two-population discrimination analysis. Section 4 shows the results of the simulation study and actual data analysis. In Section 5, the systematic analysis based on the statistical procedure is organized to investigate the development of infantile language acquisition, and the results and
discussion are provided. Finally, in Section 6 we conclude the paper and mention future research.

2. SPEECH DATA

The infant speech database (Amano et al., 2003) contains longitudinal recordings of five Japanese infants and their parents. The infants were born and raised in the Tokyo area. The parents spoke standard Japanese. A digital audio tape recorder was used for recording their natural daily conversations in a quiet room in their house. No particular task was assigned. Recording sessions were held for about 1-4 hours per month and continued for five years, and the database was developed by converting the recordings into speech files. From the speech file, a transcription file, a property tag file, a time record file, and an F0 file were made and saved in the database. Other precise information on the database is available in the previous report (Amano et al. 2003).

The number of utterances by infants and parents varied each month from twenty to several hundred. Figure 1 shows an example of an F0 pattern that includes voiced and unvoiced sounds. The vertical axis indicates the logarithmic F0 values divided by the average of F0 due to the reduction of the differences among estimated F0 values of the infant, mother and father. The horizontal axis indicates milliseconds. The unvoiced F0 values indicate missing observations as shown in Figure 1.

3. METHODS AND THEORY

3.1 REGRESSION TREND MODELS WITH MISSING DATA

To describe the F0 data having missing observations, we introduce a class of time series regression model \( \{y_t\} \) with missing points on a time period \([t_1, t_2]\). Suppose that \( \{y_t\} \) is generated by

\[
y_t = T(t) + u_t, \quad t \in \{1, 2, \ldots, t_1\} \cup \{t_2 + 1, t_2 + 2, \ldots, n\},
\]

where \( T(t) = \beta_0 + \beta_1 t + \cdots + \beta_{p-1} t^{p-1} \) with an unknown coefficient vector \( \beta \in \mathbb{R}^p \), and \( \{u_t\} \) is a stationary process. Here, note that the trend function \( T(t) \) can be defined on the whole interval \([1, n]\). First, we make the following assumptions for \( t_1, t_2 \) and the disturbance process \( \{u_t\} \).

Assumption 1.

(i) \( t_1 = \lceil \alpha n \rceil \), \( 0 < \alpha < 1 \), \( t_2 = \lceil (1 - \gamma) n \rceil \), \( 0 < \gamma < 1 \) and \( \alpha + \gamma < 1 \), where \( \lceil \cdot \rceil \) is Gauss’s symbol.

(ii) \( \{u_t\} \) is a stationary process defined by

\[
u_t = \sum_{j=0}^{\infty} \rho_j e_{t-j},
\]

where \( \{e_t\} \)’s are i.i.d. \( (0, \sigma^2) \), and the coefficients \( \rho_j \) satisfy \( \sum_{j=0}^{\infty} |\rho_j| < \infty \).

From Assumption 1, we can see that \( \{u_t\} \) is a zero-mean stationary process with spectral density of

\[
f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \rho_j e^{-j\lambda} \right|^2.
\]

Suppose that the observed stretch \( \{y_t; t = 1, \ldots, t_1, t_2 + 1, \ldots, n\} \) is available. Write
\[ Y = \begin{pmatrix} y_1, y_2, \cdots, y_i, y_{i+1}, y_{i+2}, \cdots, y_n \end{pmatrix} = \begin{pmatrix} Y_1, Y_2 \end{pmatrix} \]

with \( Y_i = \begin{pmatrix} y_i, y_2, \cdots, y_i \end{pmatrix} \) and \( Y_2 = \begin{pmatrix} y_{i+1}, y_{i+2}, \cdots, y_n \end{pmatrix} \).

Also, introduce

\[
Z_i = \begin{pmatrix} 1, 1, 1, \cdots, 1 \\ 1, 2, 2^2, \cdots, 2^{p-1} \\ \vdots \\ 1, t_i, t_i^2, \cdots, t_i^{p-1} \end{pmatrix} = \begin{pmatrix} z_i^{(1)}, z_i^{(2)}, \cdots, z_i^{(p)} \end{pmatrix}, \text{(say)},
\]

\[
Z_2 = \begin{pmatrix} 1, (t_2 + 1), (t_2 + 1)^2, \cdots, (t_2 + 1)^{p-1} \\ 1, (t_2 + 2), (t_2 + 2)^2, \cdots, (t_2 + 2)^{p-1} \\ \vdots \\ 1, (t_2 + t), (t_2 + t)^2, \cdots, (t_2 + t)^{p-1} \\ 1, n, n^2, \cdots, n^{p-1} \end{pmatrix} = \begin{pmatrix} z_2^{(1)}, z_2^{(2)}, \cdots, z_2^{(p)} \end{pmatrix}, \text{(say)},
\]

and set

\[
Z = \begin{pmatrix} Z_i \\ Z_2 \end{pmatrix}.
\]

Then, the least-squares estimator for \( \beta \) is given by

\[
\hat{\beta}_{\text{LSE}} = (Z'Z)^{-1}Z'Y = \left\{ Z_i'Y_i + Z_2'Y_2 \right\}^{-1}\{Z_i'Y_i + Z_2'Y_2\}.
\]

(3.4)

The trend function \( T(t) \) is estimated by

\[
\hat{T}(t) = (1, t, t^2, \cdots, t^{p-1})\hat{\beta}_{\text{LSE}},
\]

(3.5)

which is defined on all time points \{1, 2, \cdots, n\}.

In the actual statistical analysis, the order \( p \) of \( T(t) \) must be inferred from the data. We selected the order \( p \) by the value that minimizes Akaike’s information criterion (AIC):

\[
\text{AIC}(p) = -2 \log (\text{maximum likelihood}) + 2 \left( \text{the number of unknown parameters} \right),
\]

(e.g. Akaike 1974).

### 3.2 DISCRIMINATION ANALYSIS VIA ESTIMATED TREND

Now we investigate the problem of classifying the time series data described by (3.1) into one of two categories:

\[
\Pi_1: \text{model (3.1) with trend function } T_1(t),
\]

\[
\Pi_2: \text{model (3.1) with trend function } T_2(t).
\]

(3.6)
Suppose that we observed new data with the trend function \( T(t) \), which is assumed to belong to \( \Pi_i \) or \( \Pi_j \). Then, classification is performed by the following divergence measure:

\[
L(T : T_j) = \sum_{t=1}^{n} \left( T(t) - T_j(t) \right)^2, \quad j = 1, 2.
\] (3.7)

Actually, the trend \( T(t) \) is unknown. Therefore, we estimate it by \( \hat{T}(t) \) as defined in (3.5). Hence, the actual classification procedure is: if \( L(\hat{T} : T_j) > L(\hat{T} : T_i) \), then we choose category \( \Pi_i \). Otherwise, we choose category \( \Pi_j \). Let the probability of misclassifying the observation from \( \Pi_i \) into \( \Pi_j \) be \( P(j | i) \). It is desirable that \( P(j | i) \) and \( j \neq i \) converge to zero as \( n \to \infty \).

**Assumption 2.**

\[
\lim_{n \to \infty} \sum_{t=1}^{n} \left( T_i(t) - T_j(t) \right)^2 = \infty.
\] (3.8)

The following theorem claims that the classification based on \( L(\hat{T} : T_j) \) \( j = 1, 2 \) has fundamental goodness, and that the discriminant function \( D = L(\hat{T} : T_j) - L(\hat{T} : T_i) \) has the asymptotic normality.

**Theorem:** Suppose that Assumptions 1 and 2 hold. Then,

(i) \( \lim_{n \to \infty} P(2 | 1) = 0, \lim_{n \to \infty} P(1 | 2) = 0. \) (3.9)

(ii) Under \( \Pi_j (j = 1, 2), \)

\[
H_j [D - E\{D | \pi_i\}] \xrightarrow{\mathcal{D}} N(0, 1) \quad (n \to \infty)
\]

where \( E\{\cdot | \pi_i\} \) is the expectation of \( \cdot \) under \( \Pi_i \), and \( \{H_j\} \) is a non-random sequence, which will be defined in the proof of theorem.

We have placed the proof in Appendix.

**Remark:** Although for simplicity we restricted ourselves to the case of two-category discrimination, extension to the case of \( k \)-categories discrimination is similar. In fact, the problem is described by the following hypotheses:

\( \Pi_i \): model (3.1) with trend function \( T_j(t) \), \( j = 1, 2, \ldots, k \).

The classification procedure is: if \( L(\hat{T} : T_j) > L(\hat{T} : T_i) \), for all \( j \neq i \), then we choose category \( \Pi_j \).

Now we explain our actual calculation procedure. First, we fix two pivotal samples, say \( \{y_1^{(01)}\} \) and \( \{y_2^{(02)}\} \).

In the manner of (3.5), we make the estimators \( \hat{T}_r(t) \) and \( \hat{T}_{\pi_i}(t) \), respectively. Now, suppose that we observe data \( \{y_j^{(r)}\}, j = 1, 2, \ldots, k \), which are different to the pivotal samples. Then, we define

\[
\hat{D}_j = \sum_{t=1}^{n} \left( \hat{T}_{\pi_i}(t) - \hat{T}_j(t) \right)^2 - \sum_{t=1}^{n} \left( \hat{T}_r(t) - \hat{T}_j(t) \right)^2,
\] (3.10)

where \( \hat{T}_j(t) \) is the estimator for \( T_j(t) \).

4. **PRELIMINARY STUDY: APPLICATION TO SIMULATED AND ACTUAL DATA**

4.1 **SIMULATION DATA**

To confirm the validity of the proposed trend model and discrimination analysis introduced in Section 3, we performed a Monte-Carlo simulation study. For the noise sequence \( u \) of (3.1), general time series models
like AR, ARMA or non-Gaussian sequence are available. As a simple example, we first tried to generate 20 samples of 200 points by polynomial regression with the following AR(1) sequence:

\[ u_t = -0.53 u_{t-1} + \varepsilon, \quad \varepsilon \sim i.i.d. N(0,0.1). \]

The dimension \( p \) of the polynomial regression is decided by generating a random number between 1 and 4, while the regression coefficients are also roughly decided by generating a random number between –1 and 1. The high-order coefficients are adjusted to be small. Through the simulation study, we used the numerical calculation software Matlab® to implement the calculation procedures and the command “randn” to generate the noise sequence. We also inserted missing parts into the data where appropriate. The generated simulation data with missing parts and estimated trend data are shown in Figures 2(a) and (b). Next, we supposed two trends as \( \hat{T}_1 \) and \( \hat{T}_2 \) (actually, we chose estimated trend of 16 and 18 ) in (3.10), and calculated the \( \hat{D}_j \) value using (3.10). By describing the dendrogram obtained by the nearest neighbor method, we confirmed that the classification results were successful (Figure 3(a)). In order to interpret the physical meaning of \( \hat{D}_j \) values, let each category shown in Figure 3(a) number from C1 to C4. The \( \hat{D}_j \) value for each datum is shown by the bar graph in Figure 3(b) and the symbols from C1 to C4 are added to corresponding areas in the figure. Rising patterns classified into C1 and C2, like pivotal sample 16, are distributed in the positive area, and falling patterns classified into C3 to C4, like pivotal sample 14, are distributed in the negative area. Figure 3(c) illustrates the histogram of \( \hat{D}_j \) values that indicate the tendency of each category. Each data is clearly classified into one of C1 (a linear straight line and a slightly curvy) and C2 (exponential-like curve) in the positive area, and C3 (a linear straight and a slightly curvy) and C4 (an exponential-like concave curve) in the negative area. The simulation study confirms that the proposed statistical procedure can be effectively applied to data and that \( \hat{D}_j \) values show the characteristics of patterns classified into different categories.

4.2. ACTUAL DATA

After confirming the validity of the proposed procedure, we first applied it to the F0 data of a 24 month-old infant and its parents from the database introduced in Section 2. The reasons that we used the 24 month-old infant data are: 1. parents do not address children older than 20 months with infant-directed speech (Inui et al. 2003) and 2. Infants of around 20 months reach a stage where they are capable of concatenating words and creating a sentence structure (Kajikawa et al. 2004).

To obtain clear and normal F0 patterns, we add some conditions to select F0 data from the database. That is, the duration of the speech is less than 10 seconds, the background noise level is low, there is no overlap with other utterance, and there is no incidental or unrelated utterances such as laughing, crying, reading, hiccups, coughing.

Although the duration of utterances was limited to less than 10 seconds, several different lengths of utterance existed throughout the conversations. We first made the histogram of utterance lengths, and took the utterances that give the maximum frequency in the histogram. The frequencies of the infant, mother and father were around 300, 600, and 200, respectively, taken from a range between 200 and 400 points. Because
the father’s utterances were fewer, we used the wider range. The F0 patterns of the mother’s utterances are shown in the left panel of Figure 4(a). The right panel of Figure 4(a) indicates the estimated trend values. When applying the trend model (3.1), the length of data for each utterance completed the number by resampling to 300 points due to calculation of (3.10). We used the Matlab® command “resample” to make the length of the data uniform.

When applying the discrimination procedure, it is convenient to standardize the range of F0 data by dividing the maximum absolute value of the data. The dendrogram and corresponding trend values are shown in Figure 4(b) using pivotal samples 3 and 22. The patterns related to flat and flat-like falling, flat and rising, flat and bell-shaped, rising, falling and wavy were categorized. Although we omit to show the dendrogram in the infant’s and the father’s cases, they were also categorized to each suitable pattern.

5. SYSTEMATIC STATISTICAL ANALYSIS TO INVESTIGATE THE DEVELOPMENT OF INFANTILE LANGUAGE ACQUISITION

Our goal requires a comparison in order to systematically investigate the development of infant language acquisition. In this section, we propose the objective pivotal trend and quantitative representation to observe the changes of the pattern characteristics and similarities for each month. Using them, we arrange an automatic analyzing procedure that includes the theory introduced in Section 3 and show the results gained by applying the systematic algorithm to actual data analysis in two families’ cases.

5.1 IMAGINAL PIVOTAL TREND APPROACH

In practice, the estimated pivotal trend depends on the actual data sample. In order to discriminate by more objective criteria, we propose the usage of imaginal trends instead of an estimated trend by pivotal samples.

Suppose and are imaginal trends corresponding to categories and , respectively. Like the case defined in (3.10), we define the following divergence measure:

\[ D_j = \sum_{t=1}^{N} \left( g(t) - \hat{T}_j(t) \right)^2 - \sum_{t=1}^{N} \left( f(t) - \hat{T}_j(t) \right)^2, \]  

As examples, we used the following functions:

- case 1: \( f(t) = -t + 0.5, \quad g(t) = t - 0.5 \quad (t = 1, \ldots, N) \)
- case 2: \( f(t) = \sin t, \quad g(t) = \cos t \quad (t = 0, \ldots, N/2\pi). \)

Figure 5 summarizes the classification results in the father’s case at the time when the infant was one month old. Both panels demonstrate mostly similarly successfully classified dendrograms except for minor difference shown in the category of “flat” patterns. This shows that the imaginal pivotal trends like these work efficiently as substitutes of estimated pivotal trends in pattern classification. Also, in analyzing the longitudinal database we treat, it is considered that such objective discriminate criteria should be employed, as we have to find common pivotal trends to compare the discrimination result for each month.
5.2 SYSTEMATIC SPEECH ANALYZING ALGORITHM

Figure 6 shows the histograms of $D_j$ values obtained from (5.1). The upper, middle and lower panels in (a)-(d) indicate the cases of mother, infant and father, respectively. The data were recorded when the infant was 1, 7, 24 and 31 months old. Examination of the frequencies of the infant’s $D_j$ reveals that the height of the bar representing high frequencies decreased over time, coming to resemble the parents’ $D_j$ as months progressed. By following $D_j$, we can know the process of similarities that develop between infant and the parents’ utterances. That is, we need to propose a quantitative criterion to evaluate $D_j$ distribution as shown in the histograms.

Then, recall Theorem (ii) introduced in Section 3. Suppose that we obtain $D_j^{(1)}, D_j^{(2)}, \ldots, D_j^{(N)}$ from the data belong to the identical cluster $\Pi_j$. For $D_j^{(k)} (k = 1, \ldots, N)$, the standardized $D_j$ approximately obeys the following normal distribution low:

$$P_x [H_x | D_j^{(k)} - E(D_j | \pi) \leq z] \rightarrow \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx.$$ \hspace{1cm} (5.2)

The sample mean $\hat{m}(j)$ and sample standard deviation $\hat{\sigma}(j)$ are

$$\hat{m}(j) = \frac{1}{N} \sum_{n=1}^{N} D_{j(n)}^{(n)}, \quad \hat{\sigma}(j) = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} (D_{j(n)}^{(n)} - \hat{m}(j))^2}.$$ \hspace{1cm} (5.3)

Furthermore, using the standardized $D_j^{(k)}$ as

$$\frac{D_j^{(k)} - \hat{m}(j)}{\hat{\sigma}(j)},$$ \hspace{1cm} (5.4)

we describe the histogram for $k = 1, \ldots, N$. To investigate the configuration similarities of the histogram, we define the following measure

$$B(1:2) = \sum_{j=1}^{J} |H_j(j) - H_2(j)|$$

where $J$ denotes the total number combining each interval of the histogram, and $H_i(j)$ describes the frequency at $j$th interval in the histogram 1’s case. Thus, for each month, we compare $B(m:i)$ between mother and infant, and $B(f:i)$ between father and infant and investigate the development of the similarities.

To automatically analyze the longitudinally recorded infant speech database, we now summarize the following algorithm:

1. Load the text file containing the information about data length for each utterance and the relevant binary file with the F0 data for each utterance.
2. Specify the range of the most common frequencies for all data lengths. Practical experience indicates that any data less than 50 ms long should be discarded, since such data rarely indicates constructive sentences, just responses.
3. Within the range specified by 2, apply the polynomial regression model with missing observations. Omit the data that has a range of missing observations exceeding 25%. Choose the order of the polynomial by AIC and estimate the trend using the coefficients estimated by the least-squares method.
4. Complete the number of the trend estimates with the resampling procedure.
5. Set the appropriate estimated pivotal trends. Standardize the range of the F0 data amplitude over all utterances and all months.
6. Based on the imaginal pivotal trends, calculate the discrimination score $\hat{D}_j$ for each trend estimate.

Using standardized $\hat{D}_j$, calculate (5.3), and (5.4) for each month.

### 5.3 APPLICATION TO TWO FAMILIES’ DATA

We applied the systematic algorithm to the data of two different families, called SA and SK (these are the infants’ initials). As for estimated pivotal trends, we took case 1 given as the example in Subsection 5.1, in which the intercept was adjusted by the maximum absolute value of data, because they corresponded to a rising and falling F0 pattern. In fact, we consider that what the parents utter at the end of a sentence and in what manner indicates that they probably made self-willed utterances rather than in the manner of complex and wavy patterns.

Figures 7(a) and (b) show the B values calculated by using the sample mean $p_{m}$, the sample standard deviation $p_{s}$ and (5.4). The dotted and solid lines indicate $B(p_{m}, p_{s})$ and $B(p_{j}, p_{s})$, respectively. The changes commonly tend to become smaller after 20 months; that is, the tendency shown in the variety of F0 patterns indicates that utterances between infant and parents became more similar. We employed the standardized trends to calculate the discrimination function in (3.10). To interpret the result, let us return the fragmentary aspects of the $\hat{D}_j$ changes as shown in Figure 6. Since the maximum absolute data indicates an extreme rising or falling pattern, the center area of the density of $\hat{D}_j$ corresponds to the complex, flat and wavy patterns. As one can see in Figures 6 (a) and (b), the infants’ $\hat{D}_j$ in the early stage of less than 10 months old distributed closer to the center area than the parents’ $\hat{D}_j$ variety, which are distributed across a wider area. These results suggest that the infants might not be able to control their own utterances in the situation where intonation at the end of the sentences is rising or falling. Figures 6 (c) and (d) show that after the infants passed 10 months of age, the variation of utterances that the infants control should increase at the latter stage and the figure of $\hat{D}_j$ become close to the parents’. In Figure 7, since we employ $\overline{D}_j$, based on the objective pivotal trends instead of the estimated pivotal trends, it is presumed that the tendency we mentioned above was directly being reflected in Figure 7. Also, from developmental psychological and linguistic points of view, we can mention the following interpretations about the development of infant language acquisitions: One-month-old infants are sensitive to all range of phonemes that construct human language (Kuhl, 1987). Around seven months, infants start to utter babbles (Oller, 1980), and the first language appears around 10-13 months (Bates et al., 1987)). After that, two-word sentences appear around 17 months (Kajikawa et al., 2004), infants’ MLU starts increasing at 20 months (Inui et al., 2003) and the infants’ vocabulary spurts around 17-19 months. In particular, lexical development in comprehension and production rapidly changes around 20-22 months (Reznik and Goldfield, 1992). Adding consideration of the parents’ side to these arguments, Amano et al. (2003) pointed out that the use of a higher
F0 is one of the characteristics of infant-directed speech from the early period until 20 months. Accordingly, our results in Figure 7 support the hypothesis that the period around 20 months is the point that language acquisition processing activity clearly changes, based on objective and systematic statistical procedures.

6. CONCLUSIONS

Using a class of time series regression models with missing data, we developed a two-category discrimination analysis and a systematic algorithm in order to investigate the development of infant speech acquisition by F0 pattern analysis. The model was described within range-merging periods, except for the missing parts, and the coefficients of the regression model were obtained by the least-squares method. Because most of the F0 patterns appeared non-stationary, the proposed discrimination analysis is very practical. It was shown that the misclassification probabilities converge to zero as the length of data tends to infinity, and that the discriminant function has asymptotic normality. The simulation study indicated that the proposed statistical procedure could obtain accurate classifications. Based on the sample moments and divergence measures between mother and infant, and father and infant, the proposed discrimination analysis proved to be an efficient method for quantitatively showing pattern similarities. The statistical procedure constructed systematic program and was applied to F0 data obtained from two different families’ conversations. Our proposed procedure showed that the changes in F0 patterns between infant and parents, and these results demonstrate that a significant change in the way the infants processed language information occurred at around two years of age. Therefore, the procedure is considered to be useful in the speech processing system that requires F0 pattern classifications.

7. APPENDIX

This section provides the proof of theorem.

Proof of Theorem

(i) Let \( \mathbf{U} = (u_1, u_2, \ldots, u_{n_1}, u_{n_1+1}, u_{n_1+2}, \ldots, u_n)' = (\mathbf{U}_1, \mathbf{U}_2)' \) with \( \mathbf{U}_1 = (u_1, u_2, \ldots, u_{n_1})' \) and \( \mathbf{U}_2 = (u_{n_1+1}, u_{n_1+2}, \ldots, u_n)' \). From (3.4), it follows that

\[
\hat{\beta}_{\text{LSE}} - \beta = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{U}
\]

\[= \left\{ \mathbf{Z}_1\mathbf{Z}_1' + \mathbf{Z}_2\mathbf{Z}_2' \right\}^{-1} \left\{ \mathbf{Z}_1\mathbf{U}_1 + \mathbf{Z}_2\mathbf{U}_2 \right\}. \tag{7.1}\]

Let \( \mathbf{S}_n \equiv \text{diag}(s_1, \ldots, s_p) \), where \( s_j^2 \equiv \mathbf{z}_j^{(0)}\mathbf{z}_j^{(0)}' + \mathbf{z}_j^{(y)}\mathbf{z}_j^{(y)} \). Note that

\[
\sum_{i=1}^{n} t_i^k = \frac{1}{k+1}n^{k+1} + O(n^k). \tag{7.2}\]

From (5.2) and Assumption 1(i), it is seen that

\[
d_j^2 = \sum_{i=1}^{k} (t_i^{j-1})^2 + \sum_{i=k+1}^{n} (t_i^{j-1})^2
\]

\[
= \frac{1}{2j-1}(t_i^{j-1})^2 + \frac{1}{2j-1}(n^{j-1})^2 + \frac{1}{2j-1}(t_i^{j-1}) + \text{lower order terms}\]
\[ \alpha^{j-1} + 1 - (1 - \gamma)^{j-1} \] + lower order terms. \hspace{1cm} (7.3)

Similarly as (5.3) we obtain
\[
\lim_{n \to \infty} \frac{z_n^{(1)}}{s_j s_k} = \frac{\sqrt{2j - 1} \sqrt{2k - 1}}{j + k - 1} \times \frac{\alpha^{j+k-1} + 1 - (1 - \gamma)^{j+k-1}}{\sqrt{\alpha^{2j-1} + 1 - (1 - \gamma)^{2j-1}} \sqrt{\alpha^{2k-1} + 1 - (1 - \gamma)^{2k-1}}}
\]
\[ = A_{jk} \text{, (say).} \hspace{1cm} (7.4) \]

Next we evaluate the asymptotic variance matrix of
\[ S_n (\hat{\beta}_{LSE} - \beta) = \left\{ S_n^{-1} \left( Z_n \beta + Z_n \right) S_n^{-1} \right\}^{-1} \times \left\{ S_n^{-1} \left( Z_n U_1 + S_n^{-1} Z_n U_2 \right) \right\} \hspace{1cm} (7.5) \]

First,
\[
V \left[ S_n (\hat{\beta}_{LSE} - \beta) \right] = \left\{ S_n^{-1} \left( Z_n \beta + Z_n \right) S_n^{-1} \right\}^{-1}
\times \left\{ S_n^{-1} Z_n \Sigma_1 Z_n S_n^{-1} + S_n^{-1} Z_n \Sigma_2 Z_n S_n^{-1} + S_n^{-1} Z_n \Sigma_3 Z_n S_n^{-1} + S_n^{-1} Z_n \Sigma_4 Z_n S_n^{-1} \right\}
\times \left\{ S_n^{-1} \left( Z_n \beta + Z_n \right) S_n^{-1} \right\}^{-1}
\]
\[ = \left\{ A_n \right\}^{-1} \left\{ B_{11}^{(n)} + B_{12}^{(n)} + B_{21}^{(n)} + B_{22}^{(n)} \right\} \left\{ A_n \right\}^{-1}, \text{ (say),} \]

where \( \Sigma_0 = E \{ U_i U_j \} \). From (5.4) we can see that \( A_n - A \) as \( n \to \infty \), where \( A = \{ A_{jk} \} \), \( p \times p \) \text{-} matrix. Writing \( \sigma(j) = E \{ u_i \} \), from (5.4) and Theorem 8 of Hannan (1970, p. 216), it is not difficult to show
\[
\lim_{n \to \infty} \left\{ B_{11}^{(n)} + B_{22}^{(n)} \right\} = 2\pi f(0)A. \hspace{1cm} (7.6) \]

Recalling Assumption 1, we see that
\[
\sum_{j \in \mathbb{N}, j \neq 0} |\sigma(j)| \to 0 \text{ as } n \to \infty, \hspace{1cm} (7.7) \]

which, together with (5.4), implies
\[
B_{12}^{(n)} \to 0, \hspace{1cm} B_{21}^{(n)} \to 0, \text{ as } n \to \infty. \hspace{1cm} (7.8) \]

Therefore,
\[
\lim_{n \to \infty} V \left[ S_n (\hat{\beta}_{LSE} - \beta) \right] = 2\pi f(0)A^{-1}. \hspace{1cm} (7.9) \]

Let us return to the discriminant problem. Under \( \Pi_1 \), it is seen that
\[
L(\hat{T} : T_2) - L(\hat{T} : T_1) = \sum_{t=1}^{p} \left[ T_2(t) - \hat{T}(t) \right]^2 - \sum_{t=1}^{p} \left[ T_1(t) - \hat{T}(t) \right]^2
\]
\[
= \sum_{t=1}^{p} \left[ T_2(t) - T_1(t) - (\hat{\beta}_{LSE} - \beta)'w_t \right]^2 - \sum_{t=1}^{p} \left[ (\hat{\beta}_{LSE} - \beta)'w_t \right]^2, \hspace{1cm} (7.10) \]

where \( w_t = (1, t, t^2, \ldots, t^{p-1}) \). In view of (5.9), we have
\[
\sum_{t=1}^{p} \left[ (\hat{\beta}_{LSE} - \beta)'w_t \right]^2 = (\hat{\beta}_{LSE} - \beta)' \sum_{t=1}^{p} w_t w_t' (\hat{\beta}_{LSE} - \beta) = O_p(1). \hspace{1cm} (7.11) \]
Hence,

\[
P(2 | 1) = P \left[ L(\hat{T} : T_1, T_2) \leq 0 \right] = P \left[ \sum_{i=1}^{n} (\hat{T}_i(t) - T_i(t))^2 - O_p(1) \times \sqrt{\sum_{i=1}^{n} (\hat{T}_i(t) - T_i(t))^2} + O_p(1) \leq 0 \right]
\]

\[\rightarrow 0 \text{ (as } n \rightarrow \infty) \quad \text{(by Assumption 2)}.
\]

Similarly we can prove

\[P(1 | 2) \rightarrow 0 \quad \text{(as } n \rightarrow \infty),\]

which completes the proof of (i).

(ii) Without loss of generality we consider the case of \( \Pi_1 \). From (5.10) it follows that

\[
D - E(D | \Pi_1) = -2 \sum_{i=1}^{n} (\hat{\beta}_{\text{LSE}} - \beta)' w_i (\hat{T}_i(t) - T_i(t))
\]

\[= -2 \sum_{i=1}^{n} (\hat{T}_i(t) - T_i(t))w_i (\hat{\beta}_{\text{LSE}} - \beta)
\]

(7.12)

In view of (7.9) and Theorem 10 of Hannan (1970, p. 221) we can see that

\[S_n (\hat{\beta}_{\text{LSE}} - \beta) \xrightarrow{d} N(0, 2\pi f(0)A^{-1})
\]

(7.13)

Write \( \mu_n = -2 \sum_{i=1}^{n} (\hat{T}_i(t) - T_i(t))w_i \), and define

\[H_n = 2\pi \mu_n S_n^{-1} (2\pi f(0))A^{-1} S_n^{-1} \mu_n^{-1/2}.
\]

(7.14)

Then, (5.12)-(5.14) yield

\[H_n [D - E(D | \Pi_1)] \xrightarrow{d} N(0, 1)
\]

under \( \Pi_1 \). □

REFERENCES


FIGURE TITLES AND LEGENDS

Figure 1. An F0 pattern of voiced and unvoiced data

Figure 2. Simulation study (a) 20 samples of generated simulation data with inserted missing parts. (b) Estimated trend data (solid line).

Figure 3. Interpretation for similarities by $D_j$ values and results of clustering. (a) Dendrogram obtained by clustering. (b) Estimated $D_j$ values and correspondence with each nearest category. (c) Histogram of $D_j$ values.

Figure 4. Estimated trend and F0 pattern classification in mother’s speech. (a) Mother’s original F0 data (left panel) and estimated trend (right panel). (b) Dendrogram and corresponding F0 patterns in mother’s data.

Figure 5. The classification results in the case of numerical expressions (applied to father’s data) (a) case 1. (b) case 2.

Figure 6 Histograms of $\hat{D}_j$ values

Figure 7 Comparing the B values in SA and SK cases. (b) SK case ($B(m:i)$: dotted line, $B(f:i)$: solid line). (a) SA case ($B(m:i)$: dotted line, $B(f:i)$: solid line)
Figure 1.
Figure 2.
Figure 3.
Figure 4.
Figure 5.
Figure 6

(a) 1 months old

(b) 7 months old

(c) 24 months old

(d) 31 months old
Figure 7