Title	文音声中の基本周波数の時間変化に含まれる個人性に関する研究
Author(s)	大野, 宏
Citation	
Issue Date	1997-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1105
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

A study on Speaker individuarity in fundamental frequency contours of sentences

Hiroshi Ohno

School of Information Science,
Japan Advanced Institute of Science and Technology

February 13, 1998

Keywords: fundamental frequency contours, speaker individuarity, Fujisaki model.

1 Introduction

This paper discusses speaker individuality in fundamental frequency contours of sentences based on analysis using the Fujisaki model and psychoacoustic experiments. The stimuli used for the experiments are synthesized using STRAIGHT [1], whose fundamental frequency contours are modified by the Fujisaki model. The experiment results indicate that (1) fundamental frequency contours of sentences have much speaker individuality, (2) especially, the base frequency F_{min} and the timing parameters (T_0, T_1 and T_2) in the frequency contour have more speaker individuality than other parameters and subjects can be divided into two groups, in which fundamental frequency height or timing of fundamental frequency dynamics affects discrimination, and (3) speaker individuality can be controlled by manipulating a few parameters including timing parameters.

2 Fujisaki model

A fundamental frequency contours $F_0(t)$ [2] as follows:

$$\begin{aligned}
\ln F_0 \;&=\; \ln F_{\mathrm{min}} + \sum_{i=1}^{I} A_{\mathrm{p}i} G_{\mathrm{p}i}(t - T_{0i}) + \sum_{j=1}^{J} A_{\mathrm{a}j}\{G_{\mathrm{a}j}(t - T_{1j}) - G_{\mathrm{a}j}(t - T_{2j})\}, \\
&G_{\mathrm{p}i}(t) = \begin{cases} \alpha_i^2 t \exp(-\alpha_i t) & (t \geq 0), \\ 0 & (t < 0) \end{cases} \\
&G_{\mathrm{a}j}(t) = \begin{cases} \min[1 - (1 + \beta_j t)\exp(-\beta_j t),\, \theta_j] & (t \geq 0), \\ 0 & (t < 0) \end{cases}
\end{aligned} \tag{1}$$

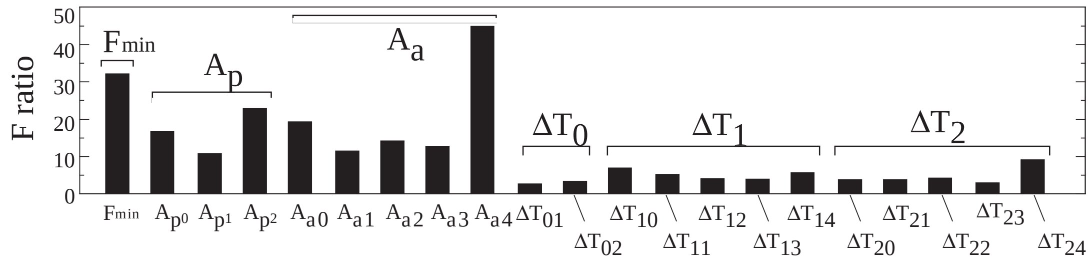

Figure 1: F ratio of each parameter

where $F_{\rm min}$: baseline value of a F_0 contour, I: number of phrase commands, J: number of accent commands,$A_{{\rm p}i}$: magnitude of the i-th phase command, $A_{{\rm a}i}$: amplitude of the j-th phase command, T_{0i}:instant of occurrence of the ith phrase command, T_{1j}: onset of the j-th accent command, T_{2j}: end of the j-th accent command, α_i: natural angular frequency of the phrase control mechanism to the i-th phrase command, β_j: natural angular frequency of the accent control mechanism to the j-th accent command, and θ_j: ceiling level of the accent component for the j-th accent command.

3 Analysis of difference in fundamental frequency contours on sentence

Speech data for all the experiments are sentences such as "aōi aōīḡā aōi yāneno uēn̄īāru"("¯ " means positions of the accent) — uttered by five male speakers.

Parameters of the Fujisaki model are estimated by minimizing the mean squared error between the extracted F_0 contour and the modeled F_0 contour on a logarithmic scale. The minimization process utilizes the analysis-by-synthesis method.

To choose some physical characteristics representing speaker individualitiy in the analyzed parameters, we calculated the F ratio (inter-speaker variation divided by averaged intra-speaker variation) for each parameter.

$$F_k = \frac{\sum_i^n \left(\bar{c_{ik}} - \frac{1}{n}\sum_i^n \bar{c_{ik}}\right)^2}{\frac{1}{N}\sum_i^n \sum_j^N (c_{ijk} - \bar{c_{ik}})^2}, \quad \left(\bar{c_{ik}} = \frac{1}{N}\sum_j^N c_{ijk}\right) \tag{2}$$

where c_{ijk} is the j-th observation of the i-th speaker for the parameter k. The larger F ratio indicates the parameter more significant for speaker classifrcation. Notes that the Δ of ΔT_{0i}, ΔT_{1j} and ΔT_{2j} indicate differences between the phase command timings and the mora boundary T_{00}.

Table 1: Psychoacouustic Experiment

Speaker	5
Subject	5
Headphone	SENNHEISER HDA 200
Headphone Amp	SANSUI AUα-907MR
Hearing level	約 76 dB (A)

Table 2: t-test of the experiment result(between synthetic speech)

stimuli sample	same	same speaker	differ speaker
O,ST	1.424	4.079	9.111
O,SF	1.585	3.654	9.199
ST,SF	1.187	0.115	0.265

$t_{0.05} = 1.960, t_{0.01} = 2.576$

4 Perception of speaker individuarity

In order to investigate fundamental frequency contours, modeled by Fujisaki model, psychoacoustic experiments used STRAIGHT speech waves with spectral and amplitude exchanged.

The types of the stimuli ars as follows:

1. O:original speech waves
2. ST:synthesized speech by STRAIGHT and TEMPO, whose spectra come from another speaker speech.
3. SF:synthesized speech by STRAIGHT and Fujisaki model, whose spectra also come from another speaker speech.

Psychoacoustic experiment was by method of parired comparison of five judge scale.

The results of t-test among three stimuli are shown in Table2 and Table3.

The experiment results indicate that (1)fundamental frequency contours of sentences have speaker individulity, and (2)fundamental frequency contours by the Fujisaki model have speaker individulity as much as those by TEMPO.

5 Shift of perception by each parameters

The psychoacoustic experiment used ABX method, the stimuli x resynthesized by exchanged a few parameter, and subjects judged whether the synsthetic speech x was closer to speaker a or speaker b.

The exchanged parameters are as follow:

1. base F_{min}

Table 3: t-test of the experiment result(between stimuli pair)

stimle	same stimuli and differ speaker	some speaker and differ speaker
ST	41.024	61.221
SF	37.722	57.52

$t_{0.05} = 1.960, t_{0.01} = 2.576$

Table 4: Parameter set

type	A	B	C	D	E	F	G	H
base	a	b	a	a	a	b	b	a
phrase	a	a	b	a	a	b	a	b
accent	a	a	a	b	a	a	b	b
timing	a	a	a	a	b	a	a	a

2. phrase A_{pi}
3. accent A_{aj}
4. timing T_{0i}, T_{1j}, T_{2j}

The exchanged parameters sets are shown in Table4.

The psychoacoustic experiment result is shown in Table5. This result is the average rate of that subjects judged speaker b.

The experiment results inducate that (1)the shift of perception affect difference of the parameters between speakers ,(2)F_{min} and the timing paramerters(T_0, T_1 and T_2) in the frequency contour have more speaker individuality than other parameters ,(3) subjects can be divided into two groups, in which fundamental frequency height or timing of fundamental frequency dynamics affects discrimination ,and (4)speaker individuality can be controlled by manipulating three parameters including timing parameters.

The results inducate that the timing parameters in the fundamental frequency contours of sentences have more speaker individuality than words. The experiment result obtain same result of the report [4], the speaker individuarity affect difference of acoustic features.

6 Conclusion

In order to investigate speaker individuality in fundamental frequency countours of sentences, parameter extraction by Fujisaki model, analysis of difference, and the psychoacoustic experiments were carried out.

The results indicate that fundamental frequency contours of sentences have speaker individuality, and timing parameter have more speaker individuality than other parameters.

Table 5: Identification rate of exchanged paramerter set

parameter set	A	B	C	D	E	F	G	H
subject1	×	△	○	△	○	◎	○	◎
subject2	×	○	○	△	○	◎	○	◎
subject3	×	○	△	△	◎	◎	○	◎
subject4	△	○	△	△	◎	○	○	○
subject5	×	△	×	△	◎	△	○	○
average	×	○	△	△	◎	○	○	◎

perceptualrate × : 0〜5 %, △ : 5〜20 %, ○ : 20〜40 %, ◎ : 40〜100 %

References

[1] H.Kawaahara："A hight quality speech analysis, modification and synthesis method STRAIGHT",J. Acoust. Soc. Jpn, pp.189-192 （1997）

[2] H. Fujisaki and K. Hirose: "Analysis of voice fundamental frequency contours for declarative sentences of Japanese", J. Acoust. Soc. Jpn. (E) **5**, 4 (1984)

[3] M. Akagi and T. Ienaga："Speaker individualities in fundamental frequency contours and its control", J. Acoust. Soc. Jpn. (E) **18**, 2 (1997)

[4] M.Hashimoto, N Higuchi：" Analysis of acoustic features affecting speaker identification", Eurospeech'95, pp.435-438 (1995)