

中国語話者と母語話者の日本語朗読音声の韻律
ー F0 モデル指令に着目した解析 ー*

○平野 宏子, 広瀬 啓吉 (東大), 河合 剛 (北大), 峯松 信明 (東大)

1 はじめに

我々は、日本語学習レベル中上級の中国語 (標準語) 話者と日本語 (東京方言) 話者が日本語文を朗読した音声の F0 について、F0 モデル [1] を用いた定量的な解析を行い、中国語話者の音声に、より高頻度でフレーズ指令、アクセント指令が生成されることを観測した [2]。中国語話者と母語話者は同一文を、各自が通常発話する速度で発話しているが、両話者の発話速度は異なる。そのため中国語話者の指令の多さは、母語話者より話速が遅いことによる影響を含みうる。本報告では、[2] とは別のデータにおいて、中国語話者と母語話者が通常速度で発話した (話速の異なる) ものと、話速が同一の範囲にあるものを用いて解析した結果を示す。

Table 1 各話者が発話した 12 文。

S#	Text
S1	昨日から頭が痛いので今日の授業は休みました。
S2	木曜日の午後にある音楽の授業が一番の楽しみだ。
S3	ビールを飲みながら皆で語り合いにぎやかな夜だった。
S4	鍋に水を入れて湯を沸かしてください。
S5	ラーメンが食べたいと思って近所の店まで歩いた。
S6	テレビを見ながら家で一人でワインを飲んだ。
S7	音楽のリズムに合わせて子ども達が元気に踊っている。
S8	何度も繰り返し練習しなければうまくすることはできません。
S9	この仕事が終わるまで帰らないでください。
S10	5 時になったら帰ってもいいです。
S11	読みたい本は借りられているので今日は借りられない。
S12	覚えることがたくさんあって全部覚えられない。

2 音声資料

2006 年冬に収録したデータセット中の、中国語話者男性 10 名、日本語母語話者男性 10 名の 12 文 (表 1) 発話を用いた。話者は、読点のないルビつきの各文を、各自が通常読む速さでポーズ位置を自由に選択して朗読した。12 文は 5 回繰り返され、各回でランダムな順序で 1 文 (1 行) ずつ提示された。母語話者はさらに 1 回、12 文を自然さを損ねない程度にできる限りゆっくりとした速度で朗読した。各音声は、16 kHz, 16 bit で収録した。収録には監督者が付き添い、読み誤りやいい淀みがあった際に再度録音を指示した。録音に入る前に朗読文と母語の対訳をつき合わせて、文意、語彙、統語構造の理解を確認し、各自で声に出して読みの練習を行った。中国語話者 (CH) の学習レベルは中上級で、北方出身、漢族、日本へ留学中の 20-30 代の標準中国語話者であり、母語話者 (JP) は 20-30 代の東京方言話者である。

3 手法

収録した音声から F0 を 10ms 毎に抽出し、音素ラベリングを行い、F0 モデルの基底周波数、フレーズ指令、アクセント指令といったパラメータを、言語情報、波形、F0 パターンを参照し、音声を聴取しながら手動で抽出した。フレーズ指令、アクセント指令の固有角周波数 α , β , アクセント成分の相対飽和値 γ はデフォルトの 3/s, 20/s, 0.9 を用いた。

その後、全発話のフレーズ指令、アクセント指令の生成数、発話時間長に関するヒストグラムを作成した (図 1)。図中、a) が通常速度で 12 文を 5 回ランダムに発話したデータによるものである。発話時間長 (右図) のヒストグラムから分かるように、全体で、CH と JP の分布範囲は有意に異なっている。そこで、各文毎に、CH と JP の分布の重なっている部分のデータを取り出して、同一範囲内の発話時間長において両者を比較することにした。まず、各 12 文 5 回発話のうち、1 文毎に、CH の最短から JP の最長の発話時間長の範囲にあるデータを取り出した。すると文によっては、CH と JP で分布範囲の重ならないものがあり、また選択された CH の話者数が限定され、データ数がかなり減り、全体を反映しないものになるこ

* Analysis of F0 Model Parameters of Japanese Prosody Produced by Chinese and Japanese Speakers. by HIRANO, Hiroko, HIROSE, Keikichi (Univ. of Tokyo), KAWAI, Goh (Hokkaido Univ.), and MINE-MATSU, Nobuaki (Univ. of Tokyo)

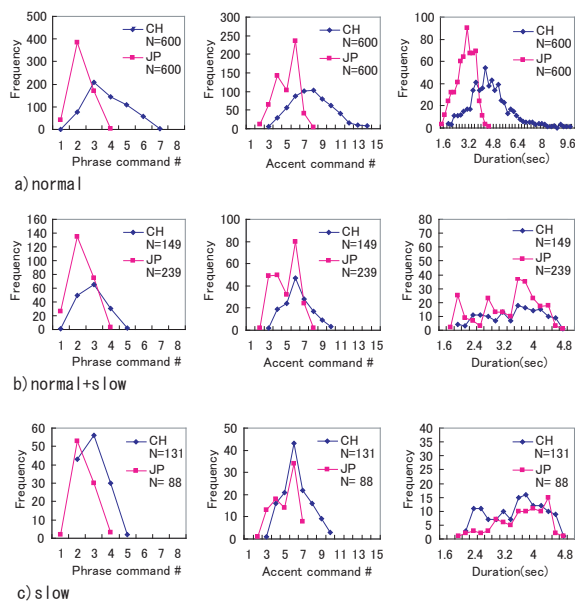


Fig. 1 中国語話者 (CH) と母語話者 (JP) の、フレーズ指令 (左図) アクセント指令 (中図) の生起数, 発話時間長 (右図) のヒストグラム.

とから, JP の通常速度のデータにゆっくり発話した 1 回分のデータを加えて選択範囲を広げた. そのヒストグラムが b) である (CH と JP の発話時間長に有意差なし). また, c) は, JP のゆっくり発話した 1 回分のデータのみと, CH の 5 回分の通常発話のデータから, 同一発話時間長の範囲にあるものを比較したものである. b) と同様に, 1 文毎に, CH の最短から JP の最長の発話時間長の範囲にあるデータを取り出した (CH と JP の発話時間長に有意差なし).

また, アクセント指令について, 各文に生成されたもののうち, 内容語, 機能語の位置に現れたものを区別して, 上記 a), b), c) の条件下で, その指令の頻度と大きさを調べた (図 2).

4 結果

図 1 のフレーズ指令 (左図) の数の平均値は, a) で CH が 3.79, JP が 2.22 と, CH が有意に多く, 同一の発話時間長の範囲を比較した b) と c) でも, b) で CH が 2.89, JP が 2.23, c) で CH が 2.93, JP が 2.39 と, CH が有意に多かった.

アクセント指令 (中図) の数の平均値は, a) で CH が 7.82, JP が 5.03 と, CH が有意に多く, 同一の発話時間長の範囲を比較した b) と c) でも, b) で CH が 6.22, JP が 4.92, c) で CH が 6.28, JP が 5.03 と, CH が有意に多かった.

なお, CH の a) と b), a) と c) の比較では, 両指令とも有意差があり, 一方, JP の a) と b), a) と c) については, 両指令ともに有意差がなかった.

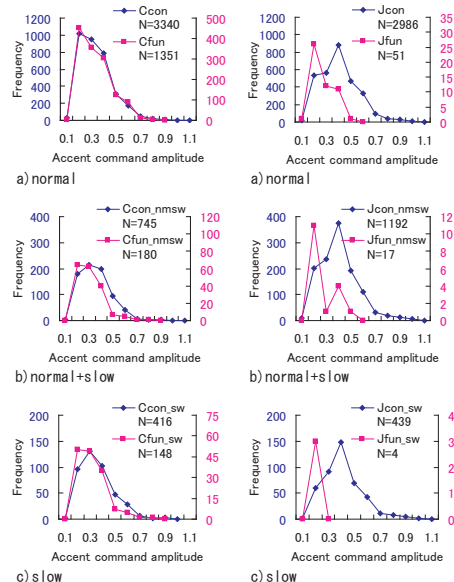


Fig. 2 内容語 (con) と機能語 (fun) のアクセント指令の大きさのヒストグラム. 左目盛が内容語の度数, 右目盛が機能語の度数.

また, 各文に生成されたアクセント指令 (図 2) は, JP では内容語に大きな指令が, 機能語に小さな指令が生起し, 機能語位置には少ないが, CH では内容語と機能語の指令の大きさに明瞭な違いが見られず, JP に比べ機能語位置の指令が多かった. 機能語に対する内容語の指令の数は, a) で CH が 2.47 倍, JP が 58.20 倍, b) で CH が 4.14 倍, JP が 70.12 倍, c) で CH が 2.80 倍, JP が 109.75 倍であった.

5 まとめ

通常速度で発話されたデータでは, フレーズ指令, アクセント指令ともに CH の生成数が JP より多かった. アクセント指令は, 文中の機能語の位置に JP よりも大きな指令が多数生成されるという特徴があった. さらに, 同一発話時間長の範囲にあるデータを調べても, 依然として同様の結果を示すことが分かった.

謝辞 本研究は, 日本学術振興会の特別研究員奨励費 (194087) の助成を得た.

参考文献

- [1] Fujisaki *et al.*, "Analysis of voice fundamental frequency contours for declarative sentences of Japanese", J. Acoust. Soci. Japan (E) 5 (4), 233-242, 1984.
- [2] Hirano *et al.*, "F0 models show Chinese speakers of Japanese insert intonational boundaries and drop pitch", Proc. Interspeech 2007, Antwerp, Belgium, 1885-1888, 2007.