

音声言語のテンポに関する一考察*

一時間構造とピッチ構造に着目して一

丸島 歩

要 旨

従来、日本語音声のテンポの速さを表す指標として、主に「一定時間におけるモーラ数」が用いられてきた。しかし、聴覚上のテンポはこの数値と一致するとは限らない。そこで、日本語母語話者を対象にした聴取実験を行った結果、速いと感じられる音声は全体的なピッチが高く、その変動も大きいという傾向が見られた。また、全体的な構造も観察したところ、ポーズを含んだ指標である「発話速度 (Speech rate)」の方が発話部分のみの速度である「調音速度 (Articulation rate)」よりも聴覚印象と相関があることが明らかになった。

キーワード

発話速度 テンポ 実験音声学 聴覚印象 聴取実験

1 はじめに

近年、いわゆるパラ言語情報に注目した音声言語を扱った研究が多くなされている。その中でも聞き手の「印象」に着目をした研究が多く見られる(山住賢司ほか 2005 など)。筆者も聞き手の印象をもとに、発話の「速度」の印象に関する研究を行ってきた。

ところで「速度」とは本来物理学的な用語であって、向きと大きさを持っておりその大きさを数値で表すことができる。つまり、一定時間における移動距離とその方向が「速度」という語の本来の意味するところである。

日本語において一般に「発話速度」と呼ばれているものは、「モーラ毎秒」「モーラ毎分」などで表される。つまり、「日本語の発話速度＝モーラ数¹/時間」と表すことができ、(向きこそ存在しないが)速度と同じように大きさが数値となって表れる²。

しかし、このような単位は発話の速度を表すのに適当であろうか。少なくとも、我々

* 本稿は、2009年11月に早稲田大学早稲田キャンパスで行われた東京音声研究会、同12月に東京大学駒場キャンパスで行なわれた東京音韻論研究会での口頭発表の内容をもとに加筆修正をしている。上記研究会でコメントやアドバイスを下さった方々に、この場を借りて厚く御礼を申し上げたい。

¹ 以前は文字数で表されることが多かったという(最上勝也 1999)が、文字数は漢字を用いるか仮名を用いるかで全く同じ音声でもその数値は異なってしまう、発話速度の表現には不適當であることは明白である。

² 「速度」は向きをもつベクトル量であるのに対し、「速さ」は向きを持たないスカラー量である。その点において、「発話速度」はむしろ「速さ」と類似している。

が日常において他人の話す速さを評価するときのような印象をも反映しているという保証はない。例えば小林聡・北澤茂良 (1996) は、発話速度を含めたプロソディー情報の聴覚印象が物理量³に必ずしも反映されない、ということを明らかにしている⁴。

筆者自身も丸島歩 (2008) などで基本周波数変動が速度印象に影響を及ぼすと主張してきた。しかし、それだけでは説明できない事例も存在した。特に、「速度」「テンポ」という時間に関わるパラ言語情報を問題にしているにも関わらず、全体としての時間構造についてはほとんど触れてこなかった。

そこで本稿では、基本周波数に加えて時間構造に焦点をあてることとする。特にポーズの存在に注目した。

2 目的

したがって本研究の目的は、発話のテンポ感によって基本周波数やポーズの特徴がどのように異なるのかを明らかにすることである。

3 方法・手順

3.1 音声資料

音声資料は、できるだけ自然談話に近いものを得ることを心がけた。実際の会話の音声を用いることが出来ればそれが最良の方法だが、複数の音声がかぶってしまうと音響解析が不可能になってしまうため、被験者が単独で話しながらもできるだけ自然に話すことができるような工夫をした。具体的には、録音の前に被験者に「最近印象に残ったこと」というテーマを提示し、テーマに基づいて自由に話してもらうようにした。その際、話す話題についてあらかじめメモを用意する事を認めた。また自然な発話を引き出すために、聞き手として筆者か第三者が頷くなどして、被験者の発話を促した。ただし、実際に過去に行なった発話を再現しようとした被験者もあり、その場合には特に働きかけはしなかった。また、さまざまなパターンの発話を得ることを目的として、①「目上の人に話すつもりで」、②「友達など、身近な人に話すつもりで」と2種類の状況を設定し、それぞれについて録音を行なった⁵。

実際に観察・実験を行なう際は、プロソディックパターンが崩れないように適当部分

³ 小林聡・北澤茂良 (*ibid.*) で言われている「物理量」が具体的に何を指しているかは明記されていない。しかし、現段階で発話速度を表す指標として一定時間内のモーラ数が広く用いられていることから考えると、それと同様か近い基準を用いていると思われる。ただし、果たしてそのような基準を「物理量」と称するに相応しいかには疑問が残る。

⁴ そのほか、聴覚レベルの発話速度を視野に入れた研究として、城生百太郎 (1999) などがある。

⁵ 録音はパーソナルコンピュータ (NEC 製 VersaPro VJ17M/ED-1) にオーディオインターフェース (CREATIVE 社製 Sound Blaster Digital Music SX) を USB 接続で介し、マイク (オーディオテクニカ社製 AT-VD4。吹かれノイズを避けるため、スポンジ状のウインドスクリーンをかぶせた) を接続して行なった。録音ソフトは Cool Edit2000、OS は Windows XP Home Edition である。サンプリングレート 44.1kHz・量子化 16bit でモノラル録音を行なった。この際、マイクと手の摩擦でノイズが発生しないように、マイクスタンドを用いた。なお、録音は静かな和室で行なった。

を切り出して用いた。さらに時間長とモーラ数が近似したものを組み合わせ、9組 (A-I) 18個の音声資料を用意した。

3.1.1 音声録音⁶

録音を行なったのは、以下の被験者である。以下に、録音時の被験者の情報を示す。

KR : 24歳女性。言語形成期 (4、5~12、3歳) は埼玉県所沢市で過ごした。

HS : 23歳女性。生後から8歳までを鹿児島県鹿児島市、8歳から9歳までを鹿児島県内之浦、9歳から10歳までを千葉県鴨川市、10歳から18歳までを鹿児島県鹿児島市で過ごした。

KA : 28歳男性。言語形成期は埼玉県所沢市で過ごした。

3.2 音声資料の音響的解析

音響解析に用いたソフトは、Praat Ver.4.6.22 (<http://www.praat.org>にて無償配布) である。同ソフトにはピッチの中央値等を算出する機能がないことから、Excel2003 SP2 (Microsoft Office社製) にデータを入力し、各統計データを算出した。解析に用いたのは、パーソナルコンピュータ (NEC製VersaPro VJ17M/ED-1) で、OSはWindows XP Home Editionである。特に本項では、ピッチと時間構造について解析した。ピッチは100Hzを基準としたsemitoneを計測し、時間構造はそれぞれポーズと発話部分を計測し、Speech RateとArticulation Rate⁷を計算した。

3.3 聴取実験

前述したように、時間長とモーラ数⁸が近似した複数の資料を組み合わせたものを用いた。各組の音声を1組ずつ聴取してもらい、「どちらが速く聞こえたか」を判断してもらった。なお、聴取はヘッドフォンによった。その上で、それぞれの基本周波数の中央値・四分位範囲、Speech Rate・Articulation Rateを計測した。なお、Speech Rateとはポーズを含めた発話速度であり、Articulation Rateは発話部分だけを計測した調音速度のことである。

⁶ 本来、実験音声学的研究を目的とする実験データは網羅的に公開されるべきであり、被験者についての情報に関しても例外ではない。父母の出身地・本人の言語生活についてなども記載すべきだが、個人情報の扱いに関する法制限により被験者名はイニシャルのみとし、年齢・性別・言語形成期の居住地以外の情報は割愛する。また、敬称は省略する。以下同様。

⁷ Speech Rate はポーズを含めた発話速度、Articulation Rate は発話部分のみで計測した調音速度のことである。

⁸ 1. において「一定時間内のモーラ数」だけが発話速度の認知を決定しているとは限らないと述べたが、ここではモーラ数を基準として解析を行なっている。現時点においては最もよく用いられているモーラを基準とするか、音節を基準にする方法が考えられるが、現在多く用いられている「一定時間内のモーラ数」という基準をできるだけ揃えることによって、聴覚的な言語テンポを決定する要因となるほかの要素を見出し得ると判断して、このような解析方法を選択した。

3.3.1 被験者

被験者は日本語母語話者の男女 14 名である。

4 結果

4.1 聴取実験

以下の図 4.1 は聴取実験の結果である。A~I 各組の結果について示してある。「サンプル平均」は一方の音声の方が速いと判断された割合である。さらに2項検定(両側)を行い、その95%区間を求めた。

このグラフから、I 以外の音声では遅速の判断が片方の音声に偏っていることがわかる。

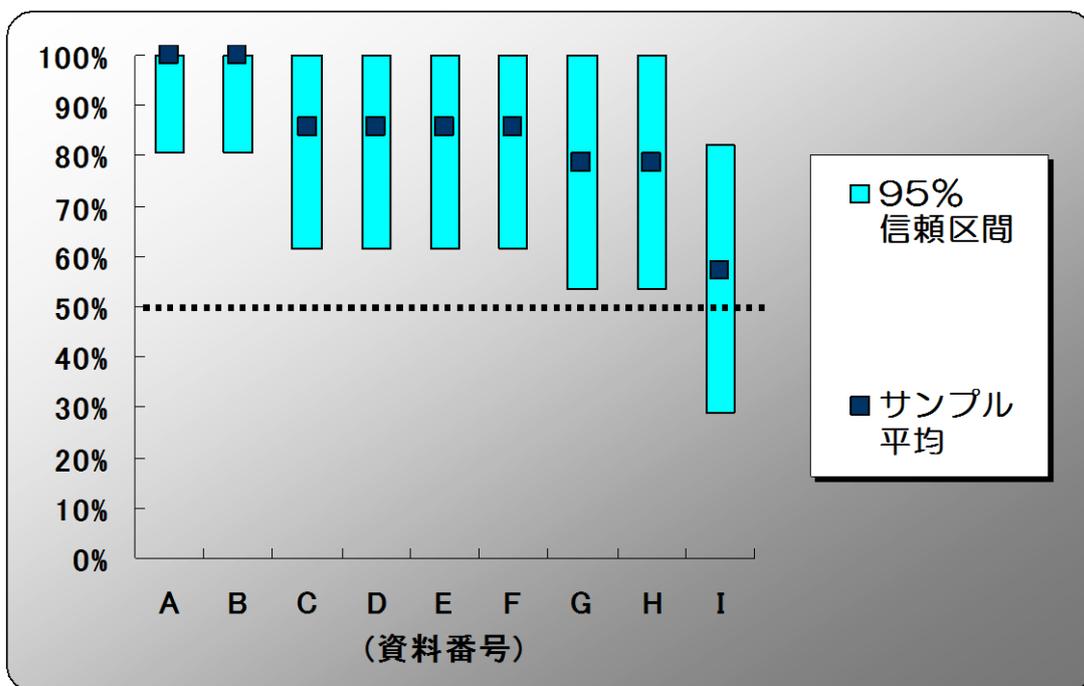


図 4.1 聴取実験結果とその統計処理結果

4.2 ピッチ情報

以下の図 4.2 に、各音声のピッチ情報を示した。各組の左側がより速いと判断された音声である。それぞれピッチの中央値とその四分位範囲について示した。さらに表 4.1 に中央値と四分位範囲それぞれと、聴取実験の結果を照らし合わせた。具体的には、より速いと判断された割合の高いほうが中央値・四分位範囲の値が大きいものを表示してある。なお、中央値・四分位範囲が大きいという判断は、大きいほうが小さいほうより10%以上値が大きい場合のみ表示してある。

以下のグラフや表から、A, B, C, E, F, H, I の組については速いと感じられたものの方が

ピッチの中央値が高く、A, B, C, E, G, H についてピッチの四分位範囲が大きいことがわかる。

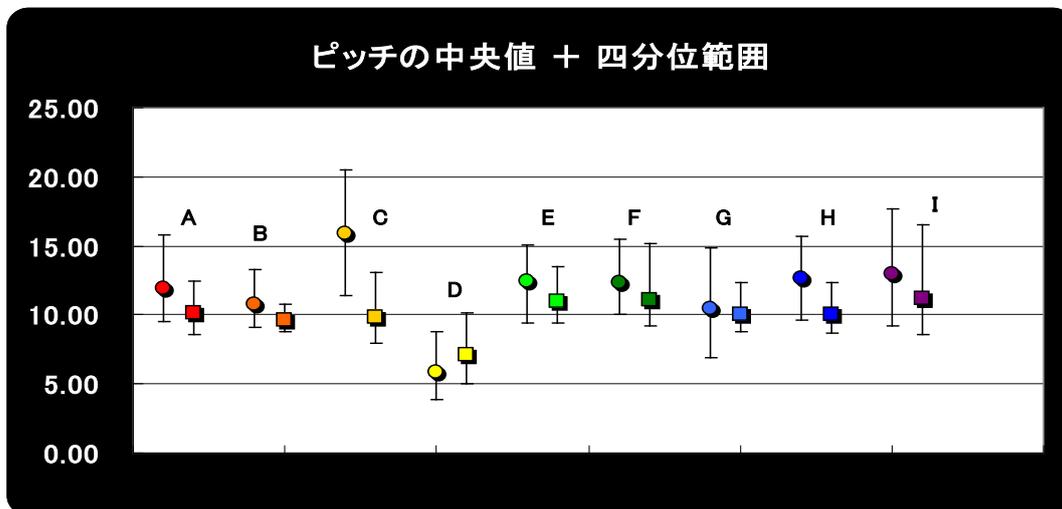


図 4.2 各音声のピッチ情報

表 4.1 聴取実験

速いと判断されたものの 中央値が大きい	A	B	C		E	F		H	I
速いと判断されたものの 四分位範囲が大きい	A	B	C		E		G	H	

4.3 Speech Rate と Articulation Rate

以下の表 4.2 は、A~J の回答の割合とそれぞれの Speech Rate、Articulation Rate である。SR1 はより速いと判断された割合が高いほうの Speech Rate、SR2 は割合が低いほうの Speech Rate である。なお、SR 比は SR1 と SR2 の比率である。AR1 はより速いと判断された割合が高いほうの Articulation Rate、AR2 は割合が低いほうの Articulation Rate である。AR 比は AR1 と AR2 の比率である。なお、比が 1.03 以上のものは青字で、0.97 以下のものは赤字で示してある。

表 4.2 Speech Rate と Articulation Rate

資料	回答の割合	SR1	SR2	SR 比	AR1	AR2	AR 比
A	100.00%	7.04	6.74	1.04	8.91	9.19	0.97
B	100.00%	6.41	6.39	1	8.47	8.77	0.97
C	85.70%	7.15	7.25	0.99	9.42	9.19	1.03
D	85.70%	5.74	5.6	1.02	7.49	8.49	0.88
E	85.70%	7	6.82	1.03	8.7	8.21	1.06
F	85.70%	9.46	9.67	0.98	10.38	10.58	0.98
G	78.60%	7.88	7.41	1.06	9.26	9.12	1.01
H	78.60%	7.8	8.24	0.95	8.99	9.59	0.94
I	57.10%	7.96	8.06	0.99	9.86	9.9	1

さらに、これらの SR 比・AR 比それぞれと、回答割合との相関係数を算出した。以下の表 4.3 に示す。なお相関係数の目安としては、0.0~0.2 が「ほとんど相関がない」、0.2~0.4 が「低い相関」、0.4~0.7 が「かなり相関がある」、0.7~1.0 が「高い相関がある」と言われている。したがって、SR 比と回答割合については「弱い相関」がある、AR 比と回答割合は逆相関になっており、なおかつ「ほとんど相関がない」ということになる。

表 4.3 SR 比・AR 比と回答割合の相関係数

SR 比と回答割合の相関係数	0.29
AR 比と回答割合の相関係数	-0.13

5 考察

5.1 考察対象

本実験では、時間長とモーラ数が近似した音声を聴き比べさせた。遅速判断に偏りが見られたのは A・B・C・D・E・F・G・H で、9 組中 8 組だった (表)。これら 8 組の音声は聴覚的なテンポ感、速度感が異なるということができよう。したがって以下の節では、これら 8 組の音声を考察の対象とする。

5.2 ピッチ動態

筆者は丸島歩 (2008) などでピッチ変動の大きい音声は速いと聴き取られやすいと主張してきた。今回の実験でも同様の傾向が見られた (表 4.1)。具体的には、今回の考察対象である 8 組中 6 組の音声は、より速いと感じられた音声のほうがそうでない音声に比べて、ピッチの四分位範囲を比較するとその値が大きくなっている。したがって、ピッチの変動が大きいほうがより発話テンポが速く聞こえるという丸島歩 (*ibid.*) を支持

する結果となっている。

ただしここで注意しておきたいのは、「ピッチの四分位範囲が広ければ広いほど、速いと判断されやすい」という傾向が見られたわけではないことである。考えられる理由としては、①速度判断は個人によって差異がみられる、②ピッチ変動以外の要因が速度感に影響を及ぼしている、という2点が挙げられるだろう。

さらに本実験の結果から、ピッチ全体の中央値が高いものの方が速く聞こえるという傾向も見られた。具体的には考察対象の8組の音声で6組について、より速いと判断された割合が高いほうがピッチの中央値が高くなっている。このことから、ピッチ動態だけでなくピッチ全体の高さも、発話の速度感に影響を及ぼす蓋然性が大きいと思われる。このことは(平均値を用いた分析であったが)小林聡・北澤茂良(1996)でも指摘されており、本研究もそれを支持する結果となった。

5.3 Speech Rate と Articulation Rate

次に、発話のテンポ感にポーズがどのように関わっているのかを考える。そのために、ポーズを加味して計測したSpeech Rateと、ポーズに左右されない数値であるArticulation Rateそれぞれが遅速判断と関わりがありそうかどうかを見た。その結果、「Speech Rate」と速度知覚には弱い相関が認められたが、「Articulation Rate」との間には相関がみられなかった。

したがって、言語テンポの知覚により大きな影響を及ぼすのは、発話部分の発話速度よりもポーズを含めた構造であることが示唆された。

ただし、ここでも注意しておかなくてはならないのは、「Speech Rate」と遅速判断との相関もあくまで「弱い」相関であるということである。このような結果になったのは、ここで算出した「Speech Rate」「Articulation Rate」があくまで「一定時間におけるモーラ数」をもとに計算されていることも一因であろうと考える。

5.4 まとめ

上記の内容をまとめると、本実験の結果から示唆されたのは①ポーズを含めた「発話速度」(Speech Rate)がいくらか発話のテンポ感の知覚に関わっていること、②しかし発話速度の影響はそれほど大きくはないこと、③ピッチ全体の高さや、変動幅がテンポ感の認知に影響を与えていること、の3点である。

6 展望

本稿では発話速度の認知が受ける影響を基本周波数との関連のみで述べた。しかし前述したとおり、筆者はこれ以外に発話速度の聴覚印象に変化をもたらす要因が存在しないとは考えていない。本稿では規定枚数の都合上触れることが出来なかったが、丸島歩(2007)ではポーズのとり方など他のプロソディー情報、文体や方言差・発話内容など

の言語レベルの問題が遅速の判断に影響を与えることを示した。

また丸島歩 (2007)、丸島歩 (2009) では音響・聴取実験だけではなく、より客観的な裏打ちを得るために基本周波数変動と発話速度認知の関わりを検証するための脳波実験も行なっている。発話速度の認知に関してより深い洞察を得るには基本周波数情報だけでなく、様々なプロソディー情報や言語情報との相関をより詳細に探っていく必要があるだろう。

さらに現時点においては、作業仮説としていわゆる「発話速度」を基礎的な単位として用いているが、加藤宏明ほか (2004) のような分節音のレベルに立った研究もわずかながら行われている。こうした研究を視野に入れながら、音声言語の時間構造の基礎的単位を模索していくべきと考える。

参考文献

- 加藤宏明・津崎実・匂坂芳典 (2004) 「音声のリズム・テンポのきこえとそのしくみ —持続時間長とタイミング処理の違い—」『文法と音声<4>』(音声文法研究会・編), くろしお出版
- 小林聡・北澤茂良 (1996) 「音声の高さ、大きさ、速さ感覚と物理関連量」『電子情報通信学会技術研究報告』 NLC96-38,SP96-69: 1-8.
- 城生佰太郎 (1999) 「現代日本語の自然音声談話のスピード」『言語』 28- 9: 44-50
- 丸島歩 (2007) 「発話速度の実験音声学的研究—聴取側の視点から—」筑波大学人文社会科学研究科修士論文
- 丸島歩 (2008) 「発話速度の知覚に関する一考察—基本周波数変動との関連性に着目して—」『言語学論叢』 オンライン創刊号
- 丸島歩 (2009) 「事象関連電位を用いた発話速度の認知実験」『言語学論叢』 特別号 城生佰太郎教授退職記念論文集
- 山住賢司・籠宮隆之・槇洋一・前川喜久雄 (2005) 「講演音声の印象評価尺度」『日本音響学会誌』 61 (6): 303-310.
- (丸島歩 筑波大学大学院生 ayumi_marushima@yahoo.co.jp)

An Observation on the Tempo of Speech Language: From a view of Duration and Pitch Construction

MARUSHIMA Ayumi

The purpose of this paper is to clarify how the Japanese speakers recognize the speech tempi.

In this experiment, some Japanese native speakers heard several pairs of Japanese natural speeches of monologues, with the same number of morae and length, and judged which sounds faster.

The results suggest that:

- (1) Rather than articulation rate, speech rate, with poses, is proportionate to the auditory tempi,
- (2) and the larger the pitch range is, the faster the informants heard the sounds.