

自然言語音を用いた発話速度の認知実験*

丸島 歩†

【要旨】筆者はこれまで、発話速度をどのように認知されるかについて脳波実験を行い、速度感とピッチ変化という異なるプロソディー情報が陰性波 N4a の潜時に反映されるという結果を得た。本稿では、全ての刺激音に自然言語音を用いてこの再現性を確認することとした。

今回の実験結果においては、必ずしも N4a 潜時に発話速度が反映されるわけではなく、最も安定して現れていた陽性波 P2 に注目すると中間的な速度である normal の刺激音がほかのものを聴いたときと潜時が異なる傾向が見られた。

この結果、速度やピッチ変化を認知する以前に、「normal か非 normal か」を二項対立的に知覚し分けている可能性が示唆された。

キーワード： 発話速度、ピッチ変動、事象関連電位

1. はじめに

1.1 問題の所在

日本語の発話速度は、しばしば一定時間内のモーラ数で表される。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)。しかし文字数を基準にすれば、漢字を用いるか仮名を用いるかで全く同じ音声でも異なる数値になり、発話速度の表現に甚だ不相当であることは明白である。

ではモーラ数を基準にすれば、発話速度を表す値として妥当なものが得られるだろうか。少なくとも、聴取側の「速度感」に対応した基準であるという保証はどこにもない。たとえば小林聡・北澤茂良(1996)では、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¹に必ずしも忠実に反映されるわけではないということが指摘されている。

1.2 先行研究

これまでの発話速度研究は、前述したような一定時間におけるモーラ数を計測する方法がとられることがほとんどである。最近のものとして、福盛貴弘 (2008) などがある。

また発話速度を聴取者の視点から観察した研究は、数は少ないものの前述した小林・北澤 (*ibid.*)などが存在する。これらの研究は聴取実験によって行なわれている。

*本稿は 2009 年 8 月 8 日に大東文化大学で開催された日本実験言語学会での口頭発表の内容に加筆修正を施したものである。発表でコメントを下さった方々に感謝の意を表したい。

†筑波大学大学院人文社会科学研究科院生

¹小林・北澤(*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。

発話速度を認知する際の脳活動に迫ったものは、城生百太郎 (1999)² のほかには見られない。

1.3 筆者によるこれまでの研究

これまで筆者は、発話速度の認知に関して聴取実験と並行して事象関連電位を用いた脳波実験による研究などを進めてきた。その結果、発話速度が速くなるほど陰性波 N4a 潜時が遅くなる傾向が見られ、この ERP コンポーネントが発話速度の遅速差を反映する可能性が示された。さらに、刺激音のピッチに変化を与えると発話速度を変化させなくとも N4a 潜時が変化することが明らかになった。このことから、ピッチ変動は発話速度の知覚に何らかの影響を与えているという可能性が示唆された³。

しかし、この時用いた刺激音は自然言語音に機械的にピッチの変化を与えた半合成音であった。心理学とは異なり、実験音声学においては自然言語音の観察が主眼となる (城生百太郎 2005) ため、刺激音の全てに自然言語音を用いた実験を行なう必要があると考えた⁴。筆者は聴取実験によっても発話速度の知覚について研究を進めているが、したがって、ピッチの変化を持たせた自然言語音の刺激音を用いてこれまでの実験結果に再現性が得られるかどうか確認することとした。

2. 目的

本稿の目的は、丸島歩 (2009) などで行なわれた一部に半合成音を用いた発話速度の認知実験に対して、全ての刺激音に自然言語音を用いた実験を行なうことでその再現性を確認するとともに、ピッチ変動が発話速度認知に影響を及ぼすかを改めて検証することである。

3. 方法

3.1 実験機器

本実験の取り込みには、筑波大学人文社会系棟 B613 音声実験室内に設置されている機器を用いた。以下の図 3-1 に機器のおおまかな配置図を示す。

²自然言語音の[papapa]と[pa]papa]のそれぞれ速いものと遅いものを刺激音として、事象関連電位を用いた脳波実験を行なっている。その結果、Pz のポイントで高次機能を反映するとされている P300 に注目すると、遅いものだけでその極性が逆転していることが明らかになった。

³詳細は丸島歩 (2009) を参照されたい。

⁴特に発話速度の認知を観察する場合、このことが重要であると考えられる。筆者は発話速度を扱った聴取実験も行っているが、実験後に被験者の方から「変わった(特徴的な)口調は速く聞こえるように思う」という旨のコメントを頂いたことがあった。このことから、音が自然であるかどうかが発話速度認知に影響を及ぼすおそれがあると考え、より自然な刺激音を用いるよう心がけた。

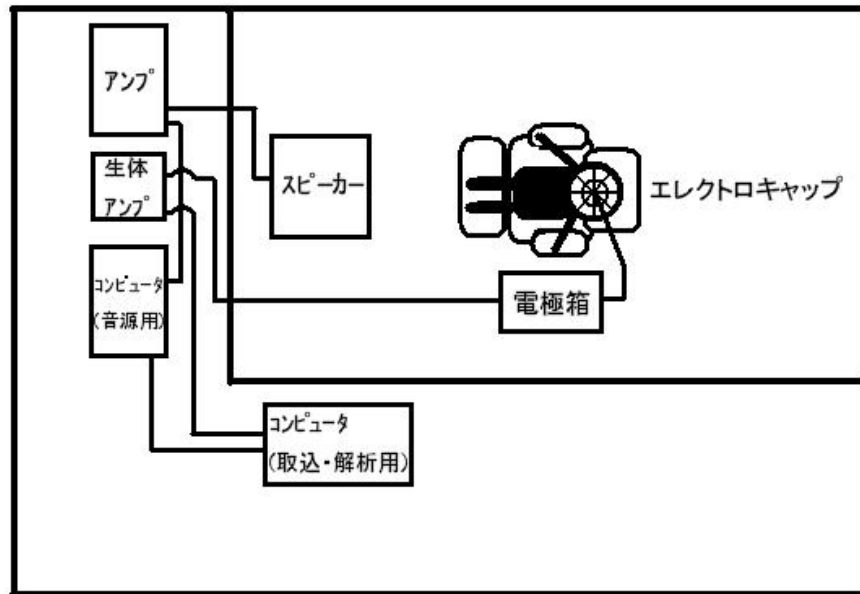


図 3.1：本実験における機器および被験者の配置

3.1.1 収録機器

被験者に装着したエレクトロ・キャップ⁵から取り込まれた脳波を、電極箱を介して生体アンプで増幅し、コンピュータで記録した。使用した機器の詳細は以下の通り。

エレクトロ・キャップ： ECI-2 (Electro-Cap International 社製)。電極配置は国際 10-20 法⁶に基づいて、F3・F4・C3・C4・P3・P4・O1・O2・F7・F8・T5・T6・Fz・Cz の 14 チャンネルを選択した (図 3-2 参照)。これを被験者に被せた上で、同社製の electro-gel を電極と頭皮の間に注入した。

基準電極：耳朶、同側耳朶法。(Electro-Cap International 社製) を使用。

電極箱： NEC ELECTRODE BOX / TYPE 6R12-2。

生体アンプ： NEC BIOTOP 6R12-2。低域遮断フィルタ 0.5Hz、広域遮断フィルタ 60Hz、感度 50 μ V/fs に設定した。

収録・記録用コンピュータ： PC98xv20 (OS は NEC98 対応の MS-DOS6.1)。用いたソフトは EPLYZER2.1 (キッセイコムテック社)。標準化 500Hz、プレトリガ-100msec、取り込み時間-100 ~ 1948msec、加算回数は 35 回に設定した。

⁵電極を頭皮上の正しい位置に固定できるよう、あらかじめ定位置に電極が取り付けられた伸縮性のある帽子。

⁶頭の前後径、左右径、周径をそれぞれ 10 等分した上で、さらに電極間の距離をそれぞれの 10% か 20% となるようにする電極の配置方法のこと(城生 伯太郎 1997:248)。

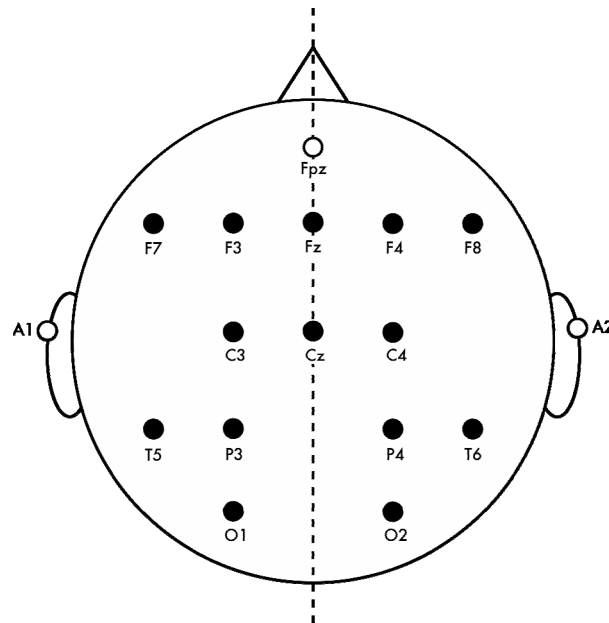


図 3.2 : 電極配置図

3.1.2 刺激発生装置

コンピュータから再生させた刺激音を、ライン・アウト端子を介してアンプ類で増幅、シールド・ルーム内のスピーカから発生させ、被験者にフリーフィールド⁷の状態聴取させた。また、同コンピュータから音声発出と同時にトリガを発し、収録用コンピュータに送ることで刺激音発出と脳波収録のタイミングを同期させた。使用した機器の詳細は以下の通り。

刺激発生用コンピュータ：IBM Vision PS/V Model 2408 (OS は Windows3.1)。用いたソフトは Winstim (榊氏作製、私家版ソフト)。単一の刺激音が 3000msec の提示間隔で繰り返し発出されるように設定した。

アンプ類：Technics Stereo Cassette Desk / RS-678U、Technics Stereo Flat Preamp / SU-9070、Technics Stereo Universal Frequency Equalizer / SH-9010E、Technics Peak/Average Meter Unit / SH-9020M、Technics Stereo Power Amplifier / SE-9060 (全て松下電器産業社製)。被験者の耳元の位置で再生音圧が 65dB⁸になるように設定した (音圧計はリオン社製・型式 NL-14 を使用)。

スピーカ：Technics Linear Phase Speaker System (松下電器産業社製)。

3.2 被験者

以下の 3 名にご協力いただいた。

SM 氏、女性、19 歳、右利き。言語形成地は茨城県水戸市。

OE 氏、女性、19 歳、右利き。言語形成地は福島県いわき市。

HY 氏、女性、20 歳、右利き。言語形成地は栃木県宇都宮市。

⁷ヘッドフォンを用いず、スピーカから流れる音声を両耳で聴取した、ということ。

⁸デシベル表示には SPL (sound pressure level) と SL (sensation level) の 2 種類のレベルが設けられている。成人男性の聞き取ることのできる最も小さい音と言われている 0.00002Pa を 0Hz とし、物理的に規定した SPL に対し、SL はヒトの感覚を基準としたレベルで、被験者の最小可聴域をもとにしている (城生百太郎 2005:458)、(中村健太郎 2001:34)。

3.3 刺激音

刺激音は、自然言語音の [papapapa]⁹を三種類の速度 (fast・normal・slow) で読み上げたものと、normalと同程度の速度で上下にピッチ変化を持たせて読み上げたもの (up・down) を用いた。したがって、刺激音は全部で5種類となる。

調音を行なったのは KR 氏、女性、24 歳。言語形成地は埼玉県所沢市。刺激音の時間長、ピッチの最高値・最低値・レンジを表 3.1 に示した。

なお、各刺激音のピッチ曲線を以下の図 3.3、3.4 に示す(ピッチについてのデータは「杉 Speech Analyzer Ver.1.07」による)。

表 3.1：刺激音の時間長

刺激音	時間長	ピッチの最低値	ピッチの最高値	ピッチレンジ
fast	590	233	266	33
normal	720	203	231	28
slow	960	217	241	27
up	730	228	429	201
down	740	211	376	165

(ms) (Hz)

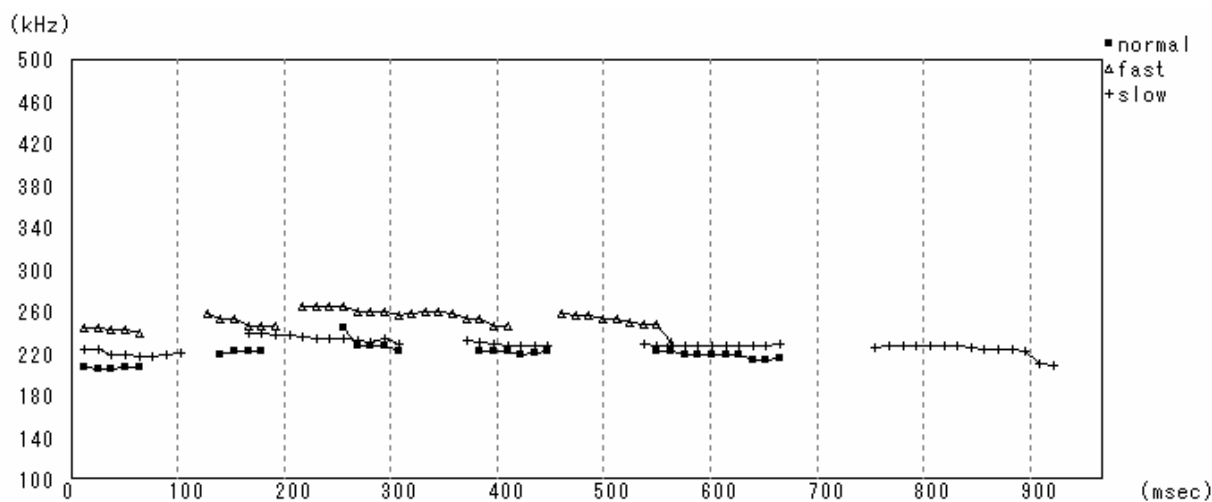


図 3.3：刺激音のピッチ曲線 (fast・normal・slow)

⁹最も脳が鋭敏に反応する子音であることがわかっている /p/(林実・寛一彦 1989) と、最も安定した母音 /a/(城生佰太郎 1997) を用いた。また、速度を知覚させるためにはできるだけ音節数を増やした方がよいと考えられるが、解析時に最大で約 1000ms を目安に観察することから、slow の刺激音が 1000ms 以下になるように 5 音節とした。

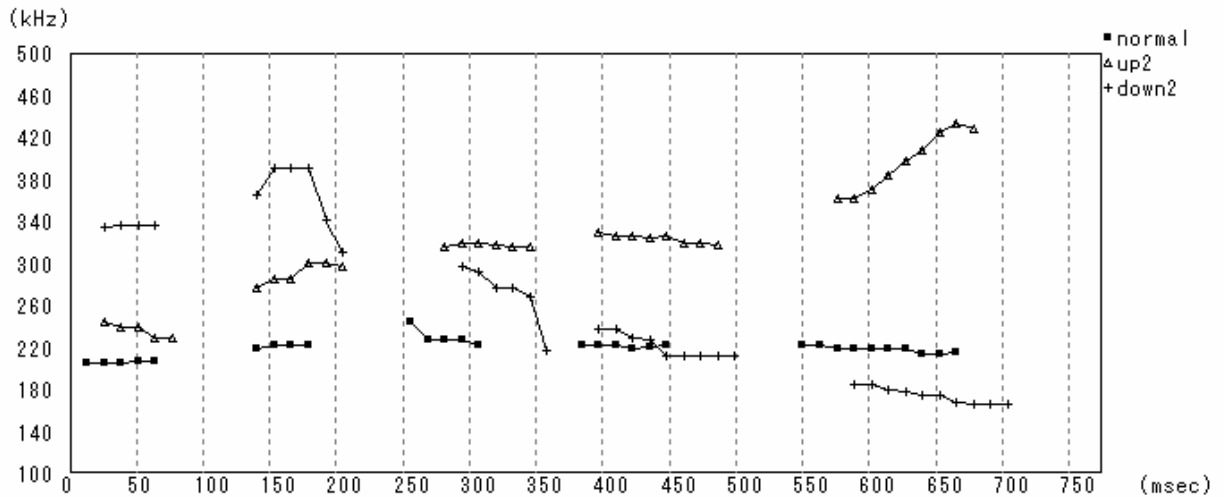


図 3.4：刺激音のピッチ曲線 (normal・up・down)

3.4 実験手順

まずは音声実験室に設置されたシールド・ルーム内にある安楽椅子に被験者を座らせ、「今から流れてくる音声を頭の中で繰り返してください」と指示し、実験中は目を半眼にしてリラックスした状態を保つように伝えた¹⁰。

各被験者に、刺激音一種類ごとに一回ずつの施行を行なった。各施行では同一の刺激を 3000 ms の間隔で 35 回提示¹¹し、35 回分の脳波を収録した。

3.5 解析方法

被験者の瞬目などによるアーチファクト¹²を除去するため、取り込みに使用したソフト (EPLYZER) を用いて RAW データ再加算¹³を行なった。

次に、解析ソフト ATAMAP の波形とトポグラフィー¹⁴を摺り合わせ、ソフト搭載のマーキング機能で陰性波・陽性波¹⁵で最も色濃くトポグラフィーが反応したところでカーソルを立てた。

3.6 データ処理方法

ピーク潜時 (PL¹⁶)をもとに情報処理を行った。

¹⁰実験前に、特に発話速度に注意を向けるような指示は行っていないが、セッションごとに被験者に刺激音の印象を聞いたところ、ほぼ全員から「fast」については「速い」、「slow」に関しては「遅い」という回答を得た。したがって、被験者はこれらの刺激音について、発話速度の差異に注目をしていただいと言えよう。それに対して、ピッチ変化を与えた「up」「down」については必ずしもピッチ変化に関する回答を得られたわけではなかった。ピッチの変化が、発話速度を含むほかのプロソディー情報の知覚に影響を及ぼしたと考えられるだろう。

¹¹各試行の冒頭は脳波の状態が安定しないため、収録コンピュータのモニターで脳波の安定が確認されてから収録を開始した。従って、実際に被験者に刺激音が提示される回数は 40 回弱である。

¹²収録中、特に電極に近い部分を動かすと、脳波に重畳してしまう筋電によるノイズのこと。

¹³収録した脳波の中からアーチファクトなどの見られないものだけを選び、再加算すること。

¹⁴脳のどの箇所がどの程度反応しているかを視覚的に見ることの出来る色分け分布図。陰性の電位を帯びた部分は赤く、陽性の電位を帯びた部分は青く表示される。電圧が強ければ強いほど色濃く表れる。

¹⁵陰性波は上向き、陽性波は下向きに表示される。

¹⁶Peak Latency の略。

4. 結果

4.1 ピーク潜時 (PL)

各実験データにおけるピーク潜時を以下の図 4.1 に示す。なお各グラフは、N1 ~ P3 の値を示してある。グラフのポイントが縦に伸びているものは、二峰性の波形¹⁷であることを表している。

さらに、これまでの実験で注目した N4a 潜時を算出し、図 4.2 に示した。なお太線で示した部分は本実験の結果であり、点線で示した部分は半合成音を交えて行った過去の実験の結果である。

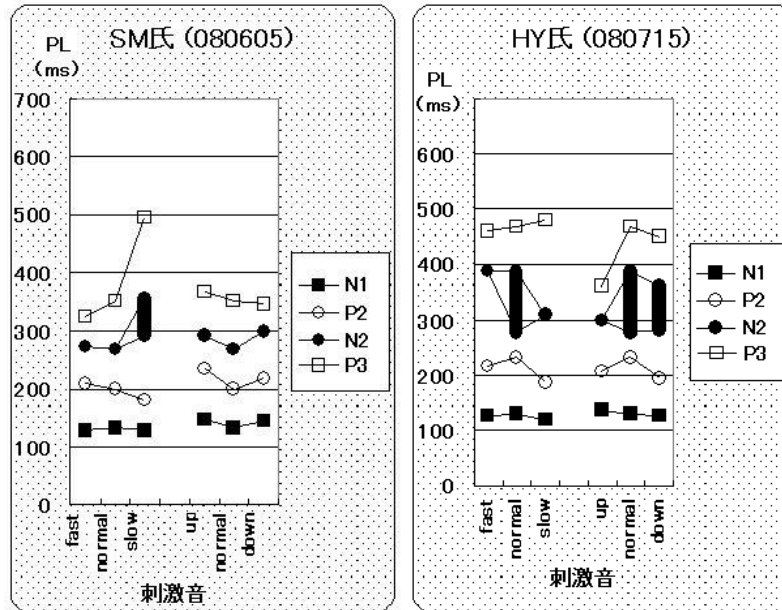


図 4.1 : N1 ~ P3 ピーク潜時

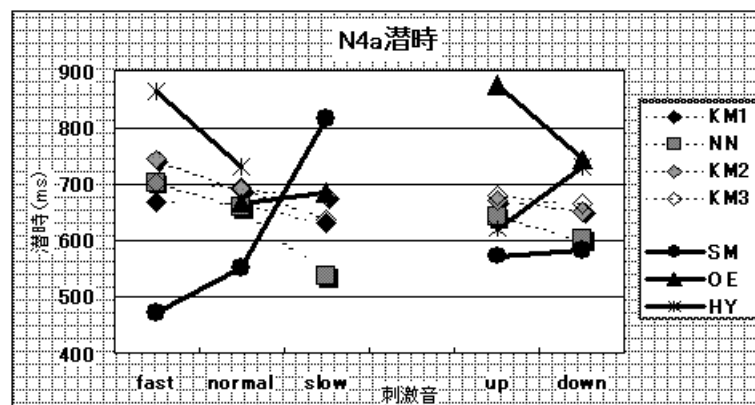


図 4.2 : N4a ピーク潜時

5. 考察

前述したように過去の実験では、発話速度が速いものほど陰性波 N4a の潜時が遅くなる傾向が見られた。しかし、今回の実験で同じような傾向が見られたものは存在しなかった(図 4.2 参照)。HY 氏については「slow」の、OE 氏については「fast」の N4a そのものが計測できなかった。

¹⁷頂点が二箇所に見える波形。一度陽性 (もしくは陰性) に振れてから、完全に陰性 (陽性) に振れずに再び陽性 (陰性) に振れる波形のこと。

た¹⁸。SM氏については刺激音の発話速度が速くなるほどN4a潜時が早くなり、これまでとはまったく逆の傾向があらわれた。このことは、SM氏がほかの被験者と異なる方略で発話速度を聴取していると考えられることもでき、これまでに観察された「N4a潜時が発話速度認知を反映している」という知見を覆すものではない。しかし城生百太郎(1997:256)にもあるとおり、刺激の入力からおおよそ700msまでが比較的安定した対応関係が見られるとされていることから、N4aを計測できなかったものがあつたことは、まったく不自然なことではないと筆者は考える。

ピッチ変動の課題のN4aについても、自然言語音を用いた今回の実験では、半合成音を用いた前回の実験結果とは異なる傾向があらわれた。具体的には、前回の実験では全ての施行においてupよりもdownのほうが潜時が早くなるという結果が得られたが、今回の実験でそのような傾向が見られたのはOEのみで、ほかの二人は全く逆の結果となった。このように今回と前回の結果に違いが見られたことは、被験者が異なることで個人差が現れたという可能性もあるが、刺激音が変わったためとも考えられる。今回このように、部分的に700msを超えるN4aというERPコンポーネントについて、以前の実験とは異なる結果が得られた。したがって、全ての刺激音に自然言語音を用いた意義があつたものとする。今回の結果を前回のそれと比較しても、N4a潜時の遅速差が示すものが何であるのかを断定することはできない。

今回の実験では、N4aの潜時が発話速度に対応するという前回の実験結果の再現性は得られなかった。そこで、本研究ではN4a以外の潜時成分に注目することとした。具体的には、被験者の3名とともに比較的安定してあらわれていた陽性波P2の潜時に注目をした。以下の図5.1に、P2潜時をまとめ直したグラフを示す。左のグラフは左からfast・normal・slowにおける潜時を示している。この三つの課題に関しては以前の実験でも全く同じ自然言語音による刺激音を用いているので、そこで得られた脳波のP2潜時も併せて考察する。中央のグラフは、今回行なった自然言語音を用いたピッチ変化をともなう刺激音の課題である。左からup・normal・slowの潜時を示してある。右側のグラフは半合成音の刺激音を用いた以前の実験結果(up・normal・slow)である。

¹⁸実験後にOE氏はfastの刺激音に驚いたと述べ、HY氏はslowの刺激音が非常に遅く感じると述べていた。これらのP4以降のピークが観察されなかったのは、このように被験者が刺激音に対して奇異に感じたことに起因する可能性があるかと筆者は考えている。

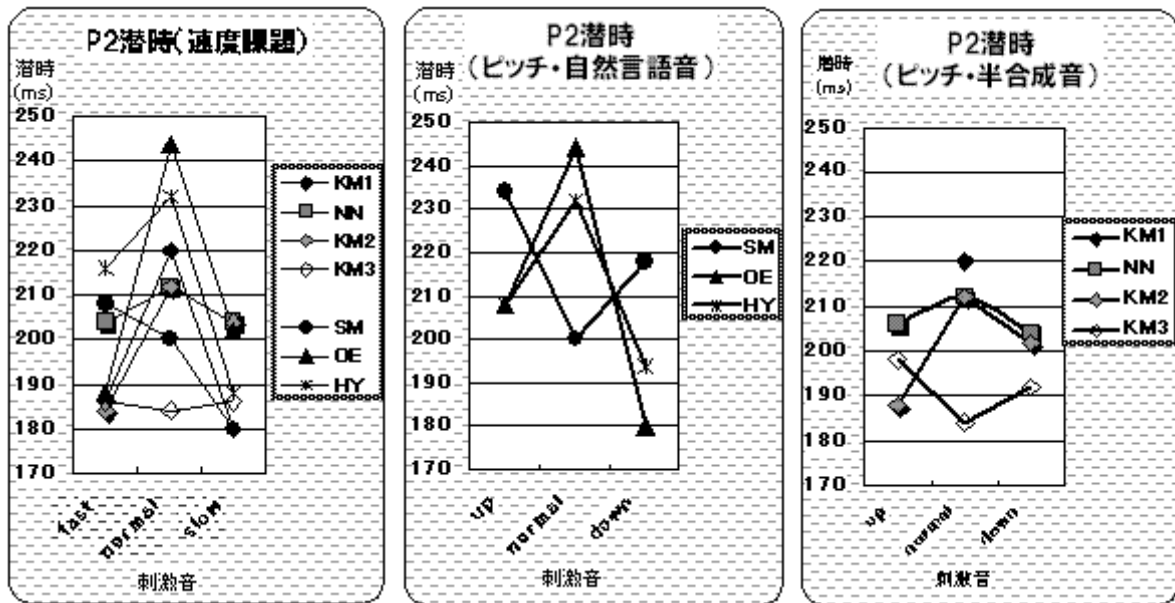


図 5.1 : P2 ピーク潜時

左(図 5.1)の速度課題のグラフを見ると、今回の実験では SM 氏以外で normal の潜時が fast・slow に対して遅くなっているのがわかる。また、ピッチ変動刺激に半合成音を用いた前回の実験においても、KM 氏(3 度目)以外は同様の傾向が見られた¹⁹。P2 以外でも normal が特徴的なものが少なくない(図 4.1 参照)。例えば HY 氏の N2 (normal がはっきりとした二峰性の波形を示している²⁰) などである。したがって、normal の音声を聴取した時は fast や slow とは異なる脳活動が行なわれていると考えられる。このことから、「“normal”と“それ以外(fast・slow)”」というように二項対立的に認知処理されていると考えられるだろう。一般に、P1・N1・P2 の段階は、刺激探知の段階と対応していると言われている。すなわち、まだ刺激が何であるのかは認識していないが、刺激が存在しているということは認識している段階である。そして、これよりも後の段階で刺激の記憶との照合が行なわれると言われている。したがって、本実験においては「刺激探知」段階の終末部にあたる P2 の時点で、その刺激が何であるのかを認識する前に「標準的なもの」と「非標準的なもの」という振り分けがなされているという蓋然性を示していると言える。

次に、中央のピッチ課題(自然言語音)のグラフを参照されたい。ここでも OE 氏と HY 氏においては、normal がほかの二者に比べて潜時が遅れている。つまりピッチ変化に差異のある音声についても、平坦に発話された「normal」とピッチ変化を与えられた「非 normal(up・down²¹)」が二項対立的に聴き分けられていると考えることができる。半合成音を用いた以前の実験(図

¹⁹ 前述したとおり、前回の実験でも速度課題には全く同じ刺激音を用いているため、ここでは今回の実験結果と併せて扱っている。

²⁰ normal のみ、ほかの刺激音とは異なる情報の処理が重畳したための二峰性波形とも考えられるが、normal の刺激音にそれにあたる音響的特徴は見出せないことから、normal の速度特有の処理が重畳したと考えられる。

²¹ down の刺激音については、ピッチ曲線を観察しても聴覚印象においても、日本語東京方言の頭高型に近いピッチ動態を示している。このことに鑑みれば、down の刺激音が被験者に「非 normal」と受け取られたとは言いは難くなる。しかし、本実験では normal 課題を初めに収録し、down 課題を up 課題の直後に収録しているため、被験者が down 課題の刺激音を「頭高型に近い音声」としてではなく「ピッチが下降している音声」として聴いた可能性が高い。いずれにせよ、この down 課題の刺激音をどう扱うべきかについては、今後の課題としたい。

5.1の右のグラフ)を確認しても、NN氏とKM氏の二度目の実験で同様の傾向が見られている²²。

ここで注目しておきたいのは、速度課題で「normalがほかの二者より潜時が遅れる」という傾向から逸脱しているSM・KM3においては、ピッチ課題においても「非normal < normal」という傾向を見ることができないということである。このことから、発話速度認知とピッチ変化の認知がある程度連動しているという可能性が考えられる。換言すれば、発話速度課題で「非normalとnormalを聞き分ける」という処理をしている被験者は、ピッチ変化でも同様の聴き方をしていると考えられる。丸島歩(2008)では聴取実験より、ピッチの変化が大きい音声はそうでないものよりも速いと感じられるということが明らかになっており、発話速度とピッチ変動の認知に関わりがあることを示唆している。したがってこの結果は、至って穏当なものであると言える。

音声言語のプロソディー²³研究において、これまでアクセントとイントネーション以外はほとんど扱われて来なかったと言っても過言ではない。本実験の結果は、発話速度の遅速差とピッチ変動という一見異なるプロソディーの知覚がそれぞれ連動していることを示唆するものであり、今までほとんど扱われて来なかったさまざまなプロソディー情報の、全容解明へ向けた重要な一歩になると考えられる。

6. 展望

これまで、発話速度とピッチ変動の関連性を観察してきたが、音圧などのほかのプロソディー特性が発話速度認知にどのように影響を及ぼすのかについては、全くの未解決である。今後新たな実験パラダイムを構築し、検証していく必要があると考えている。

【参考文献】

- 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. 電子情報通信学会
- 城生佰太郎(1997)『実験音声学研究』勉誠社
- 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』第28巻第9号:44-50. 大修館
- 城生佰太郎(2005)『日本音声学研究』平成16年度科学研究費補助金による助成出版、勉誠出版
- 中村健太郎(2001)『図解雑学 音のしくみ』ナツメ社
- 林実・箕一彦(1989)「音素・音節検出実験に基づく音声知覚の基本単位の検討」『第5回日本生体磁気学会論文集』3-1:135-136. 日本生体磁気学会
- 福盛貴弘(2008)「ニュース番組におけるアナウンサー・キャスターの発話速度」『大東文化大学外国語学部創立三十五周年記念論文集』:191-209. 大東文化大学外国語学部
- 丸島歩(2008)「発話速度の認知に関する一考察 基本周波数との関連性に注目して」『言語学論叢(オンライン版)』創刊号:70-85. 筑波大学一般・応用言語学研究室
- 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』:147-159. 筑波大学一般・応用言語学研究室

²²KM氏の一回目の実験(KM1)については、up・downの収録をおこなっていない。

²³「プロソディー」は現在アクセントとイントネーションのみの意で用いられることが多いが、ここでは発話速度や声質などを含めた広義のプロソディーを指す。

丸島歩

最上勝也 (1999) 「ニュース報道の読みの速さとその計測法」『言語』第 28 卷 第 9 号 : 40-43. 大修館

Brain Waves Experiment

about the Perception of Speech Rate

Ayumi MARUSHIMA[†]

The purpose of this study is to discuss the effect of speech rate in the cognitive process of speech sound. To serve this purpose, we observed the features of ERP elicited by listening to nonwords with various speech rates and different pitch patterns. Three kinds of speech rate were used: fast, normal, and slow. Two kinds of pitch patterns were also chosen: up and down. Three Japanese participated in this experiment.

As a result, when the subjects listened to nonwords at the normal speech rate, the peak latency of the P2 tended to advance in comparison to nonwords at fast or slow speech rates. Moreover, when the subjects listened to nonword with a flat pitch pattern (normal), the peak latency of the P2 also tended to advance in comparison to nonwords with an up or down pitch pattern.

*[†]Doctoral Program in Literature and Linguistics
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8571, Japan
E-mail: ayumi_marushima@yahoo.co.jp*