



Title	音声自然降下の聴覚情景的分析
Author(s)	松井, 理直
Citation	大阪大学言語文化学. 1993, 2, p. 117-124
Version Type	VoR
URL	https://hdl.handle.net/11094/78182
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

音声自然降下の聴覚情景的分析*

松井 理直**

This study consists of two experiments, which were carried out to investigate how fundamental frequency declination is perceived by speakers of Tokyo Japanese. In these experiments, the effects of declination to the perceptual boundary of the pitch accent and “*the ghost accent*” on devoiced sound were observed. The results are the following;

1. The perceptual boundary of pitch accent falls down to lower frequency by declination.
2. The range of this downslide of the perceptual boundary almost equals the magnitude of declination. This result suggests that the effect of declination may be normalized on the *primitive process* of the general auditory systems when word accents are recognized.
3. The effects of declination can be accounted for from the viewpoint of the “*Auditory Scene Analysis*”. One declination makes a “*stream*” and the perceptual boundary of word accent exists on this stream.

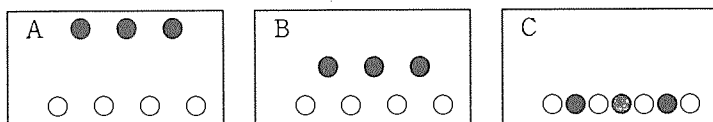
1 はじめに

まず、次の図を見ていただきたい。図Aは、白丸の列と黒丸の列が別々に並んでいるように見えるだろう。ところが、図Bになると白丸と黒丸が交互に並んでいる1つの山型の列が見えるし、図Cになると両者が一直線に並んで見える。こうした傾向は、視覚のみならず、聴覚でも観察される。図Aにあたる聴覚現象は、

*Auditory Scene Analysis of Fo Declination (Michinao MATSUI)

**言語文化研究科博士後期課程

例えば pitch が非常に離れた音が交互に鳴っているような場合で、高音の列と低音の列が別々に聞こえてくる。図Bに当たる現象は音楽でいうトレモロなどであり、図Cは同じ高さの音が連続的に鳴っている場合である。



A. Bregman (1991) は、こうした知覚現象を、“*sequential stream*”, “*grouping*”, “*segregation*”, “*capture*” といった概念によって統一的に捉え、その知覚過程を “*primitive process*” と “*schema based process*” によって説明を試みた。視覚にせよ聴覚にせよ、*primitive process* において、図Aでは *segregation* により *stream* が2つ形成され、図B, Cでは *grouping* により1つの *stream* が形成される。なお、図Aの場合、聞き方により、1つの *stream* を成すこともある。これは、注意という *schema based process* によって、*stream* を再体制化できることを示している。

本研究は、音声自然降下という現象が音調知覚に及ぼす影響についていくつかの実験を行い、その結果から、*stream* を形成する一般的聴覚システムが、音調知覚において、どのように働いているのかという点を予備実験的に考察したものである。従って、この自然降下に関する定義や、実験手法に関しては、今後再考の余地があり、あくまで研究の中間報告という形になる。

2 実験1. 東京方言アクセントの知覚実験

2.1 先行研究

Ogura, Wang & Wooters (1989) は、『雨／鉛』などの同定・弁別実験により、日本語の高さアクセントの知覚様式が範疇的知覚に従うことを示した。さらに、松井 (1992) は、音声自然降下が、範疇的知覚様式そのものには影響を与えないが、その知覚境界をより低い周波数域にシフトさせる効果を持つ（自然降下による知覚境界のシフト幅と自然降下の割合がほぼ一致する）ことを示した。

ここで、知覚境界がシフトする理由として、次の2つの可能性が考えられる。まず、高さアクセントの知覚が *baseline* (藤崎モデルというフレーズ成分) からの *pitch range* に依存している可能性である。『雨／鉛』の知覚境界を決定する

baseline からの pitch range が一定であるなら、自然降下によって baseline が徐々に下がるにつれ、知覚境界も低域にシフトすることになる。2 番目の説明としては、高さアクセントの知覚境界が、stream 上に存在するような位置に決定されるという可能性が考えられる。自然降下のない条件に比べ、自然降下が付随する条件では、右下がりの stream が形成されるため、stream 上に存在する知覚境界も必然的に低い周波数域にシフトすることになる。

本実験は、これらの仮説の妥当性を検証したものである。もし baseline の仮説が正しいのなら、他の条件に関わらず、自然降下により知覚境界は低域にシフトするであろう。逆に、stream の仮説が正しいのなら、前後の環境が同じ stream を形成しない場合には、知覚境界のシフトは起こらないであろう。

2.2 実験方法

東京都出身の男性 11 名、女性 9 名 (19 ~ 27 歳) を被験者に実験を行った。実験方法は以下の通りである。まず、音声刺激には『大粒のあめ (雨／飴)』という文章を用いた。刺激音の条件は、『大粒の』の部分に関して、次の 4 種を定めた。

1. 条件 A : *initial lowering* (IL) はあるが (おおつぶの)、自然降下のない音声。
2. 条件 B : IL があり、かつ自然降下を伴う音声。
3. 条件 C : IL がなく (おおつぶの)、かつ自然降下のない音声。
4. 条件 D : IL はないが、自然降下は付随する音声。

全ての刺激音声で、『大粒の』の『の』の pitch を 154 Hz に固定し、『あめ』の部分は 120.9 ~ 170.2 Hz まで、 $\frac{1}{14}$ octave ずつ 7 段階で、互いに独立に変化させた。自然降下は、拍内変動を防ぐために、モーラ毎に段階的に変化させた。1 モーラごとの自然降下の割合は『あめ』の部分における音高変化の割合と同じく $\frac{1}{14}$ octave とした。また、各モーラにおける定常部の持続時間は 200msec に保った。音声の大きさについては、調整法により、各モーラにおける音声の loudness が 75dB の pink noise とほぼ同一になるよう intensity を操作した。

実験手順としては、刺激音をヘッドフォンから両耳にモノラル呈示し、同定実験および ABX 法を用いた弁別実験により『雨／飴』を判断させた。各刺激音声は最終的に 10 回ずつランダムに呈示した。

2.3 実験結果および考察

音声自然降下および IL の有無に関わらず、全ての条件で、範疇的知覚が保たれた。これは、先行実験の結果と同一であり、日本語の高さアクセントが範疇的知覚に従うことを確認したものである (Fig.1-1 参照)。

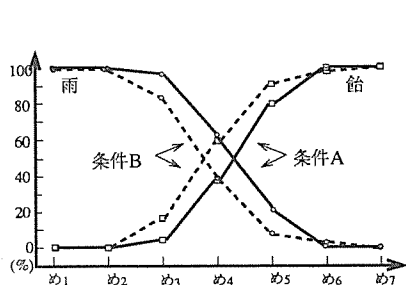


Fig.1-1 『あ3—めx』における反応様式

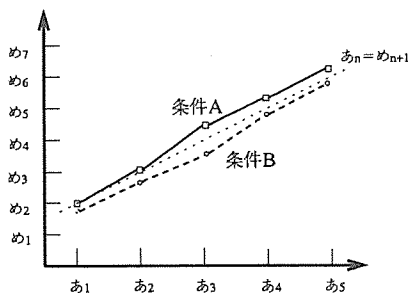


Fig.1-2 各刺激音における知覚境界

次に、その知覚境界についてであるが、自然降下の起こらない条件においては、全ての場合で、『あめ』の「あ」の音高に対し、「め」の音高がほぼ等しくなるところに知覚境界が存在した。これに対し、自然降下の起こっている条件では、基本的に、「あ」の音高よりも「め」の音高が低くなるところに知覚境界が存在した (Fig.1-2 参照)。従って、自然降下によって、知覚境界は低い周波数域にシフトすることがわかる。このことは、アクセントパターンを知覚する際に、音声自然降下の影響が、なんらかの形で修正され、除去されている可能性を示唆している。また、このシフト幅は、「あ」の音が stream の中に capture された場合（「あ」の音高が 147Hz 付近）に最大になり、ほぼ自然降下の割合と等しくなった (Fig.1-1 参照)。従って、知覚境界の位置は、baseline からの pitch range に関係しているというより、stream によって決定されていると考える方が、妥当であると思われる。

また、この結果から、『雨／飴』の弁別が、あくまでも第1モーラと第2モーラとの関係のみで決定されており、『大粒の』の「の」と『あめ』の「あ」の音高関係はそれほど影響していないことが示された。言い換えると、東京方言のアクセント知覚においては、核の位置が非常に大きな役割を果たしていることが分かる。

3 実験2. 無声拍のアクセントの知覚実験

3.1 先行研究

杉藤 (1982) は、『はしと (「し」が無声化を起こす。以下同様)』のように無声拍が起こる場合のアクセント型の知覚は、第3拍の「と」の音調が重要であると述べている。この場合、『はしと (箸と)』、『はしと (橋と)』、『はしと (端と)』の3通りが考えられるわけであるが、「と」が平板音調の場合は、『はしと (端と)』に知覚される。これに対し、第3拍が下降音調の場合で、第1拍が第3拍より高ければ、『はしと (箸と)』に、第1拍が第3拍の開始より低ければ、『はしと (橋と)』に知覚される。本節では、自然降下が伴う文発話においても、先行研究とはほぼ同様のことがいえるかどうか、また、自然降下が前述までの実験と同様の影響を及ぼすかという点に関して、実験を行った。

3.2 実験方法

被験者は、実験1と同一である。音声刺激には『赤色のはしと』という文章を用いた。刺激音の条件は、2ページ後の Fig.2-1 の通りである (自然降下の条件が変化)。その他の条件は、上記の実験とはほぼ同一である。ただし、無声拍「し」については、母音部が有声化している音を元に、調整法により loudness を 36dB の pink noise と同一にし、さらに 240Hz 以下の成分の係数を小さくすることにより作成した。また、自然降下の割合および刺激音の変動 step は、1 モーラにつき約 $\frac{1}{25}$ octave である。

知覚実験の手順は、以下の方法で行った。刺激音は、ヘッドフォンからモノラルで両耳に呈示し、同定判断を用いた弁別実験により『橋／端』を判断させた。今回の実験で、『箸』の判断を除去したのは、刺激音に拍内変動が発生していないためである。この点は、今後解決しなければならないものである。ただし、被験者が完全に『箸』と聞いた場合には、その旨を報告させた。各刺激音声は最終的に5回ずつランダムに呈示した。

3.3 実験結果および考察

『橋／端』の弁別には、無声拍『し』の音高はほとんど無関係であり、後続する『と』の音高が手がかりになっていることを確認した。これは、先行研究の結果と

ほぼ一致する。ただし、知覚様式は、『雨／鉛』の実験でみられたような範疇的知覚ほど明白な結果は得られず、最も極端な刺激音に対しても、100% の反応率は得られなかった。これは、「はし」を二者選択にしたことや、本来は起こるべき拍内変動がないためと思われる。

次に、『橋／端』の知覚境界に関して述べる。自然降下のない刺激音における知覚境界は、『は（しと）』の音高に対し、『（はし）と』の音高が約 $\frac{0.17}{25} \text{ octave}$ 高いところに位置する (Fig.2-2 参照)。

これに対し、自然降下の付随する刺激音では、『は』の音高に対し、『と』の音高が約 $\frac{1.52}{25} \text{ octave}$ 低いところに位置する。従って、知覚境界は、自然降下によって $\frac{1.7}{25} \text{ octave}$ 高い周波数域にシフトしたことになる。ちなみに、stream の仮定に基づく、本実験の場合、知覚境界は、自然降下により $\frac{2}{25} \text{ octave}$ 低い位置（即ち2モーラ分）にシフトすることになるが、実験結果はこの予測に近い値である。従って、この結果はアクセントの知覚境界が stream 上にくることを裏付けている。

同時に、この stream の仮説から、本来基本周波数を持たないはずの無声拍上にアクセントの有無を聴く理由も説明できる。『はしと』の第1モーラと第3モーラが同一 stream 上にあるときには、無声拍「し」の音がこの stream に capture されてしまうため、全てのモーラと同一の pitch を持っているかのように知覚され、『はしと（端と）』の判断が行われると考えられる。これに対し、『はしと』の第1モーラと第3モーラが同一 stream 上にない時には、第2モーラの音が stream に入るか入らないかは、一意に決定できない。基本的には、基本周波数のない弱い音である無声拍の第2モーラは、先行の stream に grouping され、第3モーラのみが segregate される。従って、第2モーラと第3モーラが違う音高であると知覚され、『はしと（橋と）』に認識されやすい。しかし、この場合は、第2モーラの音が stream に入るとは限らないため、『はしと（箸と）』の意味判断も行われるはずである。Fig.2-2 において、『端と』の判断に関しては、100% の確率になる刺激音があるが、『橋と』と『箸と』の判断が共存しているのは、第2モーラが stream に入るか入らないかが、primitive process において、決定不可能であるからだと考えられる。おそらく『はしと』のアクセントパターンに関する schema based な知識が判断に影響しており、もし、『はしと（橋と）』というアクセントパターンを知らなかったとすると、第2モーラの音高が high であるとは、聞かないものと思われる。現在、この問題に関しては、ミニチュア人工言語を用いた実験を行っている。

なお、今回の2つの実験では、自然下降が段階的についているが、これを連続的に付随させ、拍内変動を生じさせると、さらに興味深い結果が得られる。この点に関しては、他の機会に述べる予定である。

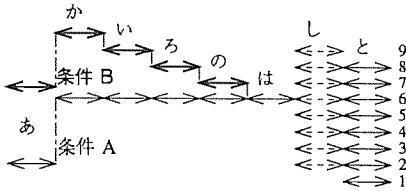


Fig.2-1 実験に用いた刺激音声

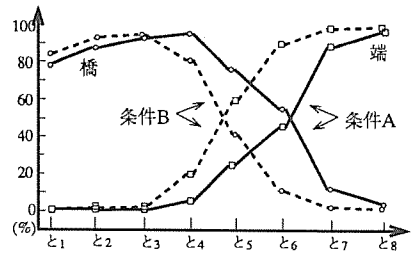


Fig.2-2 被験者の反応（「し」を丸め込み）

4 総合論議

東京方言、無声拍のアクセント知覚には、それぞれ決まった音響の手がかりが存在することを確認した。また、これらの知覚境界の存在する位置は、音声自然降下を stream として捉えることにより予測可能であり、さらにアクセント成分の抽出は、stream に対する grouping, segregation によって可能であることが示された。これらの結果から、アクセントを知覚する過程において、音声自然降下は primitive process において normalize され、同時にこの過程でアクセント成分の抽出もある程度行われ、最終的に、schema based process でアクセントの照合がなされ、単語認識がなされていると予想される。

また、こうした stream に基づく音調知覚の面から、Pierrehumbert & Beckman モデル (PB モデル) の特性を考えることもできそうである。PB モデルの特徴の1つは、音高の散在表示と、それらの直線補間による音高指定にあるが、これらを知覚の面から捉え直すと、音高の散在表示は特徴抽出に相当し、直線補間は stream の形成に相当すると考えられる。今後は、本文中で指摘した実験上の不備な箇所とともに、こうした点を検討していくつもりである。

参考文献

Bregman, A. (1991) : *Auditory Scene Analysis*. The MIT Press

藤崎 博也 (1989) : 日本語の音調の分析とモデル化. 杉藤 美代子編 『講座
日本語と日本語教育 第2巻 日本語の音声・音韻 (上)』 明治書院

松井 理直 (1992) : アクセント知覚における音声自然降下の影響に関する一考
察. 日本音響学会平成4年度秋期大会講演論文集 2-9-19

Ogura, M., Wang, W.S-Y., & Wooters. C. (1989) : W.S-Y.Wang 編 『発話
の基礎』 p84-85. 研究社

Pierrehumbert, J.B. & Beckman, M.E. (1988) : *Japanese Tone Structure*.
The MIT Press

杉藤 美代子 (1981) : 『日本語アクセントの研究』 三省堂