振幅包絡情報に基づく韻律知覚
Prosody Perception based upon Amplitude Envelope Information

同志社大学　工学部

Faculty of Engineering, Doshisha University

力丸　裕

Hiroshi Riquimaroux

Abstract: Recent works on perception of noise-vocoded speech sound (NVSS) have revealed that amplitude envelope information is very important for speech perception when spectral information is not sufficiently available. Basically, the fundamental frequency information is not available and formant peaks cannot not be identified in NVSS. However, we can even recognize intonation and distinguish male voice from female voice in NVSS without the fundamental frequency. More, prosodic information can be perceived from NVSS without the fundamental frequency, such as accent, Also, melody can be created from lyrics once lyrics are intelligible. In the present study, findings from fMRI measurement are introduced to univasively show neural activities in the central nervous system. The present data indicate that creating intelligibility for NVSS requires activities in various sites in the central nervous system which are not ordinarily used for speech recognition. Applications of the present work include an innovative speech processor and a training system for hearing impaired people.

Key words: noise-vocoded speech sounds, fundamental frequency, amplitude envelope, central plasticity, speech processor

1. はじめに

音声知覚には、周波数情報が重要な手がかりであると以前から報告されてきたが、音声知覚には、振幅包絡情報も大きくかかわる事実が、劣化雑音音声を用いた研究で明らかになってきた(Shannon et al., 1995; 小畑と力丸, 1999)。我々のグループの実験によって、基本周波数(F0)の存在しない劣化雑音音声でも、日本語文章の知覚、さらには、アクセントの知覚が可能であることもわかった。本研究では、劣化雑音音声を用いることにより、振幅包絡情報に基づいた韻律情報の知覚を検討した。

2. 劣化雑音音声知覚

2.1 劣化雑音音声の合成と文章の知覚

音声知覚には、周波数情報が重要な手がかりであると従来から報告されてきた。ところが、周波数情報が欠落し振幅包絡が残存する状況を人工的に創りだすと、振幅包絡情報が音声知覚に大きく貢献する事実が、劣化雑音音声を用いた研究で明らかになってきた[1), 2)]。すなわち、音声信号を３つまたは４つのバンドノイズに置換し、振幅包絡情報をもとのまま残し、周波数情報を極端に減らした劣化雑音音声は(noise-vocoded speech sound, **Figs. 1, 2**)、はじめて聴くとほとんど了解できないが、短期間の訓練で初めて聴く文章でも80%以上の了解度を得るようになることがわかった(**Fig. 3**)。文章より単語、通常単語より無意味単語の順番に正答率は下がるが、母音の認識においては, すべて75%であることも確かめられている。はじめて聴いたときに了解度が殆どないという事実から、通常の神経経路を用いても、劣化雑音音声の音韻性は知覚できないと推察できる。短期間の訓練後に了解できるようになる結果は、通常とは異なった経路を用いて劣化雑音音声が処理され、既存の経路統合されている可能性を示唆している。

2.2 アクセントの知覚

通常、日本語イントネーションの知覚は、基本周波数(F0)の時間的変化を用いて行われる。ところが、劣化雑音音声においては、F0が存在しないにもかかわらず、日本語のイントネーションの知覚・弁別が可能であることもわかった。例えば、「箸」(h**A**shi)と「橋」

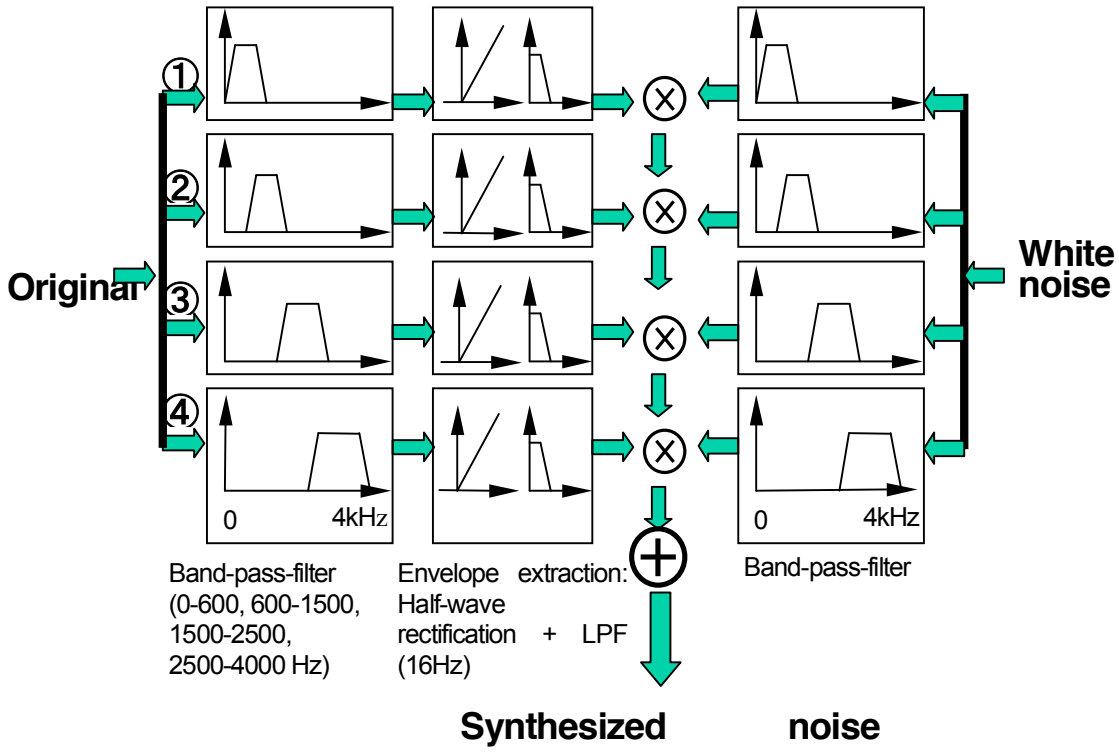

Fig. 1 Four-band speech processor. This figure schematically shows how to produce 4-band noise-vocoded speech sounds.

(hashI)である。「箸」ではF0が第1モーラから第２モーラにかけて下降し「橋」ではF0が第1モーラから第2モーラにかけて平坦もしくは上昇する(Fig. 4)。また、振幅包絡もF0と同じ方向に変化する。ここに、周波数情報を欠落させず、振幅包絡を周波数変化と矛盾した方向に変化させても、すなわち「箸」の振幅包絡を「橋」の振幅包絡に置換しても、「箸」は「橋」とは知覚されない。ところが、音声信号を劣化雑音に置換してしまうと、振幅包絡の時間的変化によって、イントネーションが決定する。換言すると、「箸」の振幅包絡を「橋」の振幅包絡に置換すると、「箸」が「橋」と知覚される。すなわち、周波数情報が欠落している場合には、脳が振幅包絡情報に基づいてイントネーションを創作していると考えら得る。したがって、韻律知覚にも振幅包絡情報がおおきな役割を果たすことが可能であることを示している[3)]。

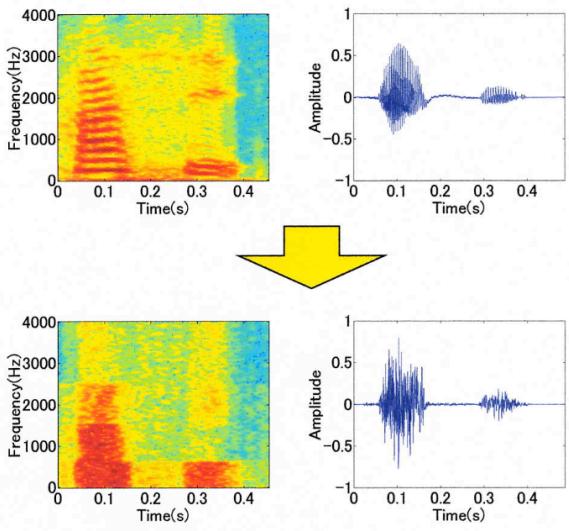

Fig. 2 Comparison between an original speech sound *"hashi"* (chopsticks) and a synthesized noise-vocoded speech sound from the original.

2.3 メロディーの知覚

通常、歌唱曲のピッチは、F0によって決定されることが知られている。ということは、メロディー知覚は、F0の時間的変化によって知覚されると考えられる。ところが、我々が用いてきた劣化雑音音声では、最も低い雑音帯域は0から600Hz までであるので、この雑音帯域に入ってしまうF0は周波数としては弁別不可能

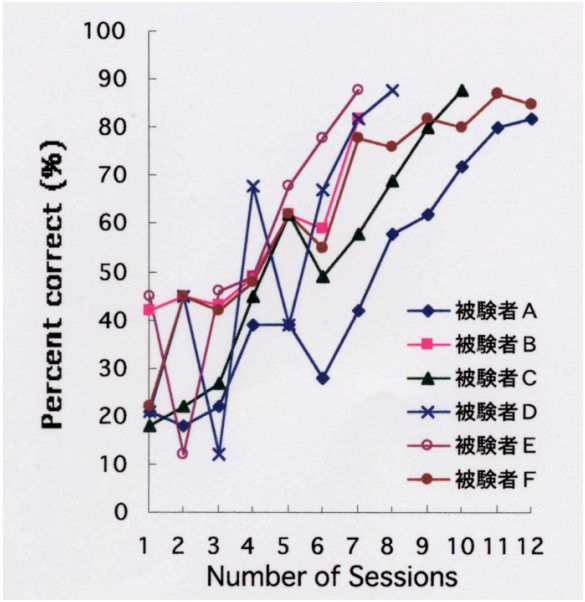

Fig. 3 Relationship between number of session and percent of correct response for 6 subjects. Percent correct exceeds 80% after 6th to 8th session.

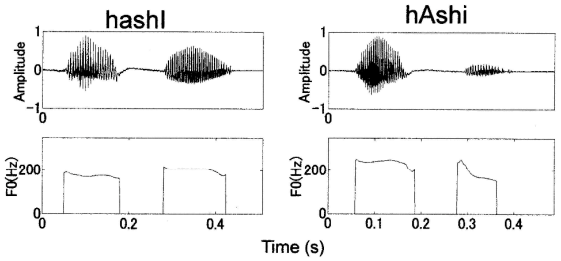

Fig. 4 F0 and amplitude envelopes of original speech sounds HASHI (chopsticks and bridge).

である(Fig. 5)。したがって、同一雑音帯域内の F0 によるメロディー情報の知覚は困難である。しかし、歌唱曲を劣化雑音に変換した知覚実験の結果、歌唱曲のメロディーを知覚することもある程度可能であることもわかった[4]。この場合、劣化雑音によって合成された歌詞が知覚されると、その歌詞に基づいて新密度の高いメロディーが脳内で創成され、そのメロディーが知覚される傾向があるのである。したがって、1番はよく知られ、2番は知られていない曲では、1番ではメロディーが知覚されるけれども、2番を聴くとメロディーが知覚されないという興味深い結果となる。

2. 4個人識別

劣化雑音音声は、帯域雑音の組み合わせであるの

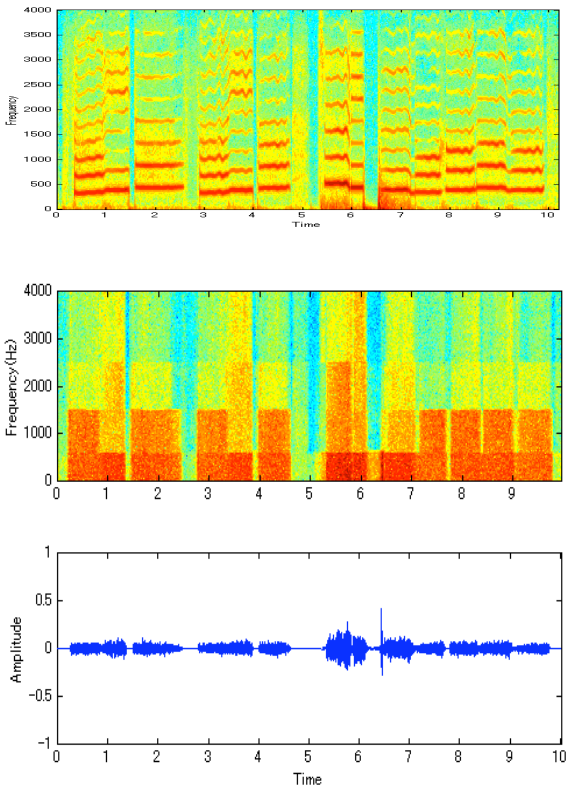

Fig. 5 Soundspectrograms of song of *"Tulips"* before and after passing through 4-band noise speech processor. Top: original sound. Middle: song replaced by 4-band noise. Bottom: temporal amplitude change in the processed song.

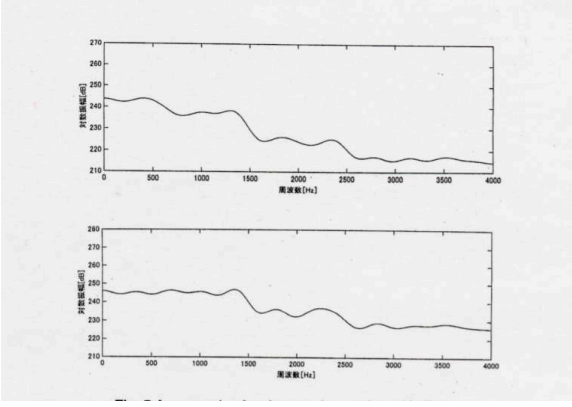

Fig. 6 An example of male speech sound used in Exp. 2 before (top) and after modifying spectral envelope by mimicking female spectral envelope (bottom).

で、F0は存在しない。また、フォルマント周波数（声道の共鳴周波数）も同定できない。したがって、劣化

雑音音声において個人を同定することは、困難である。しかし、実験の結果、男女声の弁別は可能である。これは、高周波帯域の雑音のパワーの違いによるものと考えられる[5]。すなわち、男声の高周波帯域の雑音パワーを上昇させると女声と知覚される(**Fig. 6**)。

3. 劣化雑音音声知覚と脳活動

通常、劣化雑音音声を聴いても音声としての明瞭性はない。すなわち、訓練なしでは、内容を全く理解できない。換言すると、通常の音声情報処理機構では周波数情報を主に用いて音声を解読するので、周波数情報が殆ど欠落した劣化雑音音声の明瞭化処理は不可能と考えられる。したがって、劣化雑音音声が明瞭化されるには、脳神経系における通常の音声信号処理に用いられない神経機構による信号処理が実行されている可能性、脳の可塑性の発現、が考えられるのである。そこで、4帯域劣化雑音を用いて聴取訓練を実施して充分な了解度が得られたと判断された後に、4帯域劣化雑音、1帯域劣化雑音、自然音声、定常雑音に対する脳賦活部位を functional MRI（fMRI）で測定する実験をおこなった結果、劣化雑音音声処理機構には、個人差が存在することがわかった(**Figs. 7a, b, e**)。また、通常の音声知覚の際に賦活される部位(**Fig. c**)に加えて、前頭葉をはじめ多くの部位に活動が見られた(**Fig. 7e**)。すなわち、劣化雑音音声知覚機構は、通常の聴覚機構にくわえて、脳の記憶、学習、発声機構など、多くの機構の機能が統合されたものであることが示唆された[6]。

4. まとめと今後の展望

以上より、劣化雑音音声知覚には、脳の多くの部位が関わり、脳の可塑性が必要であると考えられる。振幅包絡情報による韻律知覚も可能であることが示唆された。fMRI 測定と心理物理学的測定の併用により、脳内韻律情報創成機序が解明され、本研究の将来的展望は大きく広がっていくと考えられる。

5. 謝辞

本研究は、科研費特定領域研究（B)「韻律に着目した音声言語情報処理の高度化」の補助を受けて行われた。fMRI 装置の操作や実験方法への助言など，多大な協力を頂いた ATR 脳活動イメージングセンタ（BAIC）の諸氏に謝意を表す.

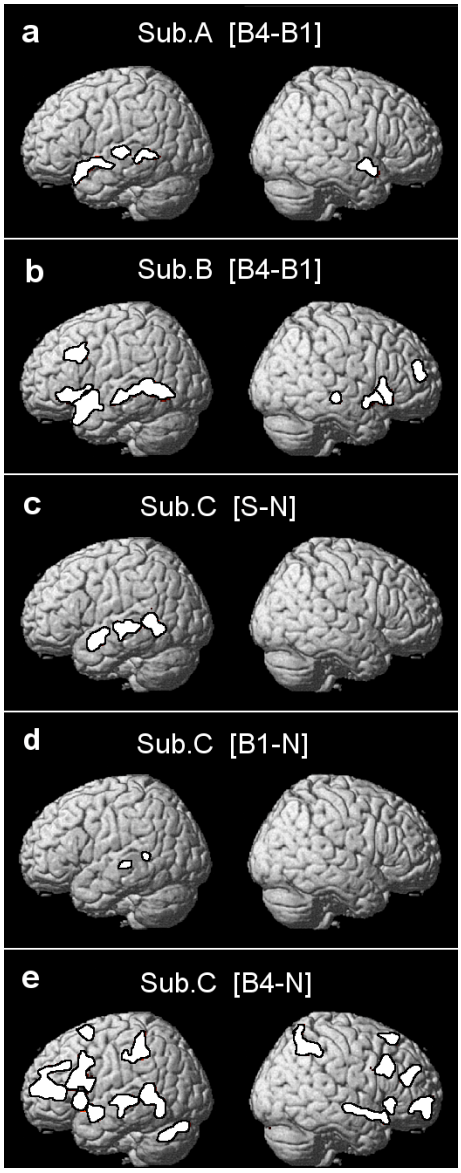

Fig. 7 Areas showing significant activities under each conditions. **a:** [B4-B1] on Sub.A. **b:** [B4-B1] on Sub.B. **c:** [S-N] on Sub.C. **d:** [B1-N] on Sub.C. **e:** [B4-B1] on Sub.C. Activations are thresholded at $p<0.05$ (corrected) and excluded the cluster that its size is less than 100.

参考文献

1) Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski and Ekelid, M.：Speech recognition with primarily temporal cues. Science 270: 303-305 (1995).
2) 小畑宜久，力丸　裕: 細時的振幅変化に着目した周波数成分劣化音声知覚の検討. 聴覚研究会資料 H-99-6 (1999).
3) 小畑宜久，力丸　裕: 帯域雑音により合成された日本語音声の了解度- 聴覚中枢神経の機能を利用したスピーチプロセッサを目指して- . 聴覚研究会資料 H-2000-3 (2000).
4) 力丸　裕: 基本周波数情報のない歌唱メロディー知覚は可能か：劣化雑音音声. 聴覚研究会資料**32**: 77-84 (2002).
5) 力丸　裕，片山貴史: 劣化雑音音声知覚はどこまで可能か？話者

弁別. 聴覚研究会資料**33**: 25-30 (2003).
6) 橘　亮輔, 力丸　裕：劣化雑音音声知覚の脳内機構：fMRI による計測. 音講論集(春): 411-412 (2004).