

## 超遅速演算を用いた実時間音声情報処理： 脳の可塑性を利用した劣化雑音音声知覚

力丸 裕

Hiroshi Riquimaroux

同志社大学工学部知識工学科

### 1. はじめに

従来の脳機能研究は、刺激となる信号に対して、脳神経系がどのように応答するかを測定する際に、静的な信号に対する静的な応答として取り組んできた。しかし、殆どの刺激は時間的に変化する動的なものであり、脳神経系の応答も時間的に変化する動的なものである。我々の脳神経系は、リアルタイムで多くのことを手際よく処理している。しかし、その処理速度（演算速度）は極めて遅く、たかだか 1 kHz のクロックに相当する程度であるので、現在のパソコンの演算速度の 300 万分の 1 程度である。聴覚や視覚などの感覚系では、リアルタイム処理に加えて、不良設定問題を効率良く解く必要がある。演算速度に依存した現在一般に行われている「認識」技術では、不良設定問題を解くことは困難である。本論文では、極端に遅いプロセッサでリアルタイム、時分割処理を実施している脳神経系に着目し、通常存在する情報が欠落しても安定した知覚を創り出すための脳神経機構の好例であるヒトの音声知覚機構のロバストネスに関して論ずる。ここで論ずる研究の方向性は、ME 分野のヒューマンインターフェイスの改良や効率の良い障害者用診断・訓練機器の開発の実現化につながる。

### 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出すと、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<sup>1), 2)</sup>。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, Fig. 1)、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。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 イントネーションの知覚

通常、日本語イントネーションの知覚は、基本周波数 (F0) の時間的な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」(hAshi) と「橋」(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。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、振幅包絡の時間的な変化によって、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<sup>3)</sup>。

#### 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 の時間的な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である。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一定程度可能であることもわかった<sup>4)</sup>。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<sup>5)</sup>。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。

### 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、劣化雑音音声の明瞭化されるには、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十分な理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. 2a, b, e)。また、通常の音声知覚の際に賦活される部位(Fig. 2c)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. 2e)。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<sup>6)</sup>

### 4. まとめと今後の展望

以上より、脳の可塑性を利用したスピーチプロセッサの可能性が充分示唆される。脳の可塑性を利用した障害者用音声知覚支援システムの実現に向けて、失語症患者、人工内耳装着患者、難聴患者を用いた聴取訓練の実現と、これらの患者での fMRI 測定が必要となる。また、訓練装置としてだけでなく、障害診断装置としても活用可能であろう。本研究の将来的展望は大きく広がっていく。

### 5. 謝辞

本研究は、文部科学省関西西文化学術研究都市地域知的クラスター創成事業「高度マンマシンインターフェイス技術群のネオカデンへの応用に関する研究」、学術フロンティア「知能情報科学とその応用」、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」、ならびに、文部科学省(旧科学技術庁)振興調製費目的達成型脳科学研究推進制度「ヒトを含む霊長類のコミュニケーション」の補助を受けて行われた。

### 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: 継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).

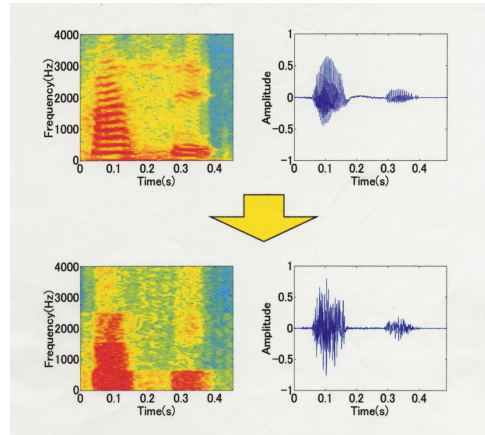


Fig. 1 Comparison between an original speech sound "hashi" (chopsticks) and a synthesized noise-vocoded speech sound from the original.

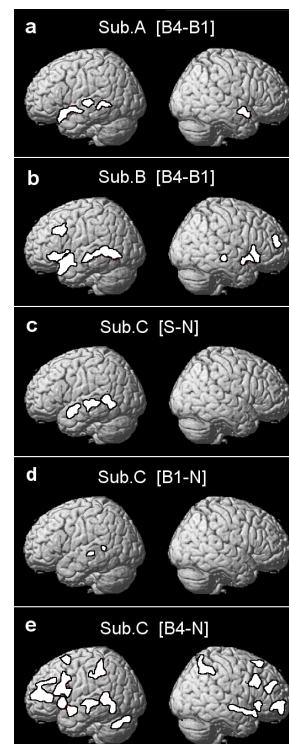


Fig. 2 Areas showing significant activities under different stimulus conditions. a: [B4-B1] on Sub.A. b: [B4-B1] on Sub.B. c: [S-N] on Sub.C. d: [B1-N] on Sub.C. e: [B4-B1] on Sub.C. Activations are thresholded at  $p < 0.05$  (corrected) and excluded the cluster that its size is less than 100.