

令和 6 年 6 月 27 日現在

機関番号：33921

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K11959

研究課題名（和文）雑音環境下の音声知覚における情報統合メカニズム

研究課題名（英文）Integration mechanisms of speech perception under noisy environments

研究代表者

小林 まおり（Kobayashi, Maori）

愛知淑徳大学・健康医療科学部・教授

研究者番号：90451632

交付決定額（研究期間全体）：（直接経費） 2,800,000円

研究成果の概要（和文）：アナウンサーと一般人の音声を用いて、音声の消失帯域幅を操作し、音韻修復現象が生じる程度を比較検討した。その結果、アナウンサー音声では250から4000 Hzの周波数帯域の情報消失した場合でも音韻修復現象が生じた。加えて、音響特徴が高周波数帯域および広周波数帯域にわたる音素の音修復率が高いことがわかった。また、250 Hz以下の周波数帯域の消失が音韻修復に及ぼす影響について検討したところ、基本周波数によって音韻修復の程度が影響されることが示唆された。従来研究では基本周波数は直接音韻の手がかりとはならないと考えられてきたが、断片的な音響特徴を統合する上で重要な役割を果たすことが示唆された。

研究成果の学術的意義や社会的意義

本研究では、雑音環境下でも頑健な音声は雑音によって知覚的に音韻が修復される音声であると考え、音韻修復現象を用いて検討した。その結果、頑健な音声では幅広い周波数帯域にわたって音声情報が消失したとしても知覚的に音韻修復されることがわかった。残余成分に着目すると、これまで音韻の直接的な手がかりと考えられてこなかった高周波数成分や基本周波数成分が音韻修復に関わっていることが示唆された。むしろ、これらの周波数成分は断片的な音韻情報を統合するための手がかりとなるのではないかと考えられる。このことは雑音下での音声メカニズムを検討する上で情報統合メカニズムの必要性を示唆するものといえる。

研究成果の概要（英文）：We examined the effects of bandwidth of speech disappearance on the phonemic restoration (PR) using professional announcers' speech and non-professional speakers' speech. As a result, it was shown that in the announcers' speech, the PR occurs even when information in the 250-4000 Hz frequency band is lost. It was found that the rates with acoustic features spanning high and wide frequency bands were higher in the announcers' speech. This suggests that elements in the high-frequency band serve as phonemic cues only when noise added. Additionally, to examine the impact of the loss of frequency bands below 250 Hz, we investigated the PR by manipulating these bandwidths. This result indicates that, although the fundamental frequency is not considered a direct cue for phonemic perception in previous studies, it plays an important role in integrating fragmentary acoustic features.

研究分野：実験心理学

キーワード：音声知覚 雑音環境下 音韻修復現象

1. 研究開始当初の背景

ヒトにとって、音声は他者とのコミュニケーションにおける重要なツールである。しかし、音声コミュニケーションは常に静かな環境で行えるとは限らない。ときにはうるさい雑音環境下でやり取りする必要もある。雑音がある状況でも相手の発言を聞き取ることは、社会生活を営むうえで必要不可欠な能力である。しかし、雑音中にヒトがどのようにして音声を知覚するのか、という問題には決定的な説は提案されていない。

雑音下での音声知覚の代表的なモデルの一つに、Glimpsingモデル(Cooke, 2006; Jones & Rosen, 1993; Buss et al., 2004など)がある。一般に、目的音と雑音が同時に存在する場合、両者の周波数が近いと聴覚末梢系での符号化の時点で両者が混ざり、目的音の検出は困難になる。ただし、音声は時間的にも周波数的にも変動するため、高エネルギーの領域が時間-スペクトル上でまばらに分布する。つまり、雑音が大きい場合であっても、時間-スペクトル上には音声信号が支配的になる領域が局所的に生じる。この局所的な領域を「Glimpsing (垣間見える)」領域と呼ぶ。Glimpsingモデルでは、まず音声を時間-スペクトル上で小さな領域に分割し、その中からglimpsing領域を信号対雑音比(SNR)に基づいて抽出する。そして、各glimpsing領域でそれぞれ音韻のパターンマッチングを行い、もっとも一致する音韻に決定するというものである。このモデルでは、glimpsing領域が音声全体に占める割合によって雑音中の音声了解度を予測できる(Cooke, 2006; Jones & Rosen, 1993; Ozmeral et al., 2012)。

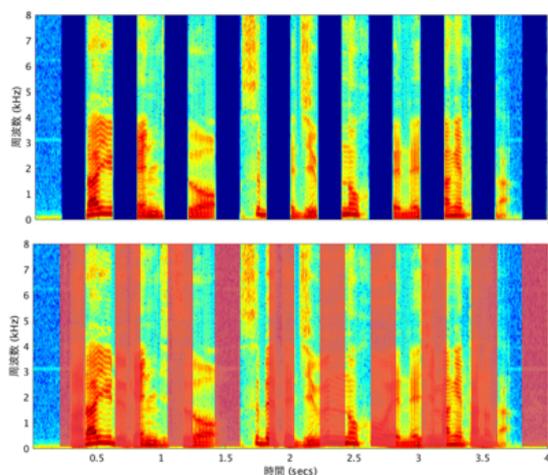


図1 音韻修復現象の例 実際には途切れている音声(上図)が、途切れている消失部分に別の音(雑音)を挿入すると(下図)、あたかも繋がっているかのように知覚される

しかし、ヒトはglimpsingのような断片的な情報をバラバラに知覚しているわけではない。このことは音韻修復現象(Miller & Liclider, 1950; Warren, 1977)が如実に示している。音韻修復とは、図1のように、実際には音の一部が消失しているにもかかわらず(図1上)、消失部分に別の音を挿入することによって、別の音の背後にあたかも音が存在しているかのように知覚される現象である(図1下)。雑音の有無で音韻の情報量は変わらないため、Glimpsingモデルでは説明できない。むしろバラバラの手がかりを統合し、消失部分を推定するメカニズムの存在を示唆している。この聴覚系の働きを考慮し、Glimpsingモデルの修正も提案されている

(Cooke, 2006; Fogerty, et al., 2018)。しかし、glimpsingを抽出する際の領域の大きさや閾値の操作にとどまっておらず、音韻修復が示すヒトの「物理的に提示された情報以上のものを知覚する」機能を積極的に考慮した音声知覚モデルには至っていない。実際、Glimpsingモデルに基づいた音声合成手法では雑音下でも頑健な音声を実現できておらず、雑音環境下の音声知覚モデルとして不十分であることを示唆している。音韻修復にみられるような適応的な聴覚情報処理過程へのモデルの修正が求められている。

2. 研究の目的

本研究では、雑音下での音声知覚のメカニズムの一端を明らかにするために、時間-スペクトル

ル的に分布した音韻の手がかりを統合する過程について検討する。音韻修復現象を用いて、聴覚における情報統合の法則（発見的規則）と発声機能の制約（音声独自の手がかり）が統合に果たす役割を明らかにする。最終的には、これらの情報統合の役割を定量化し、Glimpsingモデルに組み込むことを目指す。

3. 研究の方法

本研究は3つのサブタスクに分けて研究を実施した。まず、発声機能の制約が音声情報の統合に及ぼす影響について検討するために、聞き取りやすさに関連する音響特徴について概観し、整理を行った。特に聴取実験を行う上で話者の性別をどの程度考慮すべきか予備検討として性別による聞き取りやすさの違いとその要因について探ることを目的として行った。また、スペクトル情報における発声機能の制約を考慮するうえで声帯由来の情報と声道由来の情報とに分類し、その影響について検討した。これらの結果を踏まえ、2つ目のタスクとして雑音下でも聞き取りやすいとされる音声と一般的な音声を比較し、一般的な音声に比べて聞き取りやすい音声では音韻修復が生じる程度が異なるか聴取実験を行った。その際、スペクトル情報の影響を操作するために、消失周波数帯域幅を変化させ、音韻の手がかりや統合するための手がかりが含まれる周波数帯域について検討した。3つ目のタスクでは、聴取実験の結果から、音韻修復が生じやすい音素とその音響特徴について分析した。聞き取りやすい音声と聞き取りにくい音声で音韻修復の程度に違いが現れた音素の音響特徴を分析することで、各音素の音響特徴と、特徴を統合するための手がかりとなる音響特徴について検証した。これらの検証を通じて、従来のGlimpsingモデルでは触れられてこなかった統合するための手がかりとしてのスペクトル成分について明らかにすることを旨とする。

4. 研究成果

本研究で得られた成果を項目ごとに以下に示す。

(1) 性別による聞き取りやすさの違いとそれに関与する音響特徴

音韻修復現象における発声機能の制約（音声独自の手がかり）について検討するために、まずこれまでの研究を振り返り、音声知覚における聞き取りやすさと音響特徴の関連性について整理した。本研究は時間-スペクトル的に分布した音韻手がかりを統合するための手がかりに着目する。一般に男声よりも女性の方が基本周波数が高く、音声のスペクトルも高周波数帯域に分布するといった発話者の性別による音声成分の違いがある。そのため、こうした発話者の性別によるスペクトル成分の違いをどの程度考慮すべきなのか、聞き取りやすさとの関連性に着目し、検証することとした。

音声の性別による聞き取りやすさについて先行研究を概観したところ、女声の方が男声よりも理解度（被験者に有意義単語を音声提示する聞き取りテストの正答率）が3~10%程度高い傾向があることがわかった。この傾向は音声を雑音とともに提示した条件で特に顕著であった。また、音声の聞き取りやすさの印象評価実験においても概ね女声の方が評価が高く、言語圏に関わらず「聞き取りやすい」という印象を持たれていることが示唆された。

一方、性別による音声の音響特徴の違いについては、上述の基本周波数、基本周波数の変動量、声質（女声は息漏れごえ、男声は嗄声）、母音空間（フォルマントスペース）の広さ、有声開始時間（VOT）、発話レベルなど多岐にわたって認められることがわかった。しかし、従来研究で聞き取りやすさの要因として多く指摘されている発話時間については、性別による違いが統一して認められているわけではないことがわかった。この研究に先立ち、申請者も日本語音声データベースを用いて性別による発話長の違いを検討したが、やはり話者の個

人差が大きく、性別による違いとしては認められなかった。

これらのスペクトル上の性別の違いを音源フィルタ理論に則り、音源情報（声帯由来の音響特徴）とフィルタ情報（声道由来の音響特徴）に分類し整理したところ、特にフィルタ情報が聞き取りやすさとの対応がよいことが見られた。たとえば、母音空間は女声のほうが広いのに対して、男声は中舌母音化しやすく、母音空間が狭い。また、子音についても、男声では破裂音の弾音化や閉鎖音の解放時の周波数が低下しやすいなど音素の音響普遍性の欠如が強いことが示唆された。実際、女声に比べ男声では異音が多いことが報告されており、各音素に対しての音響特徴がはっきりと示されていることが「聞き取りやすさ」に強く関連していることが改めて示唆された。

一方、声帯情報については聞き取りやすさとの関連性が不明な音響特徴が多いことがわかった。基本周波数の高さは音声の聞き取りやすさとは概ね無関連と結論づけられているが、基本周波数の変動量（F0幅）については議論の余地が残されていることが窺えた。また、近年の研究により声帯に由来する情報が各音素の音響特徴を補完する情報を提供することが示唆されており、声帯に由来する情報を検証する必要性もあることがわかった。

以上、多くの知覚実験から女声の方が男声よりも聞き取りやすいことが示されており、話者の性別について配慮する必要があることがわかった。ただし、多くの先行研究においても女声の平均よりも了解性が高い男声話者がいることが報告されており、性別に由来する違いよりも話し方などの違いが聞き取りやすさには大きく影響することを確認した。しかし、音声一般の特徴を考えるうえで発話者の性別については偏らず検証する必要があるものと思われる。またこれら先行研究の概観から、本研究における音声の発話上の制約として声道由来の情報について着目していくこととした。

(2) 音韻修復現象における補完のされやすさの違い

雑音があることによって音韻が知覚的に修復されやすい（補完されやすい）音声にどのような音響特徴があるのか検証するために、実際に音韻修復現象が生じる条件下で音声による違いが生じるか、知覚実験を行った。音声刺激として、雑音下でも聞き取りやすいとされるプロのアナウンサー男女各2名とが発話した音声と一般成人男女各2名が発話した音声を用いて、音韻修復のされやすさについて比較、検討した。

実験では聴覚健全な成人男声10名が聴取者として参加した。音声刺激としては3～5モーラの単語を帯域ノイズと共に提示した（S/Nは固定）。音声刺激のスペクトル情報がどの程度消失したとしても知覚的に補完されるのか、その補完されやすさに違いは生じるのか検討するために、音声刺激にノッチフィルタをかけ消失帯域幅を操作した。条件としてはノッチフィルタの下限遮断周波数を250 Hzと固定し、上限遮断周波数を1～4 kHzの幅で1 kHzずつ変化させ、4条件とした。共に提示するノイズにはバンドパスフィルタをかけ、音声の消失部分にノイズの帯域が重複するように操作を行なった。

実験の結果を図2に示す。どちらの話者タイプにおいても、消失帯域幅が広がると了解度は低下する（図2A, B）。しかし、アナウンサーが発話した音声では、ノイズが付加された条件の方がされていない条件よりも了解度が高く、音韻修復現象が生じることが示された（図2B）。また、アナウンサー音声と一般人の音声でノイズなし条件の了解度を比較したところ、有意差は認められなかった（図2C）。このことから、音声の属性に関わらず消失帯域が増加すれば音声の聞き取りは同様に悪くなるものの、アナウンサー条件ではノイズが挿入されることで知覚的に補完されやすいことが示唆された。アナウンサー音声は雑音下でも聞き取りやすいことが報告されているが、ノイズがあった状況下で聴覚的補完が生じやすい音

声であることもその一助となっている可能性が示された。また、ノッチフィルタの上限周波数が4 kHz条件下でも音韻修復現象が認められることから、250～4000 Hz帯域という音声のスペクトル情報としては大部分を消失している場合においても、アナウンサー音声では雑音によって知覚的に修復されることが示された。

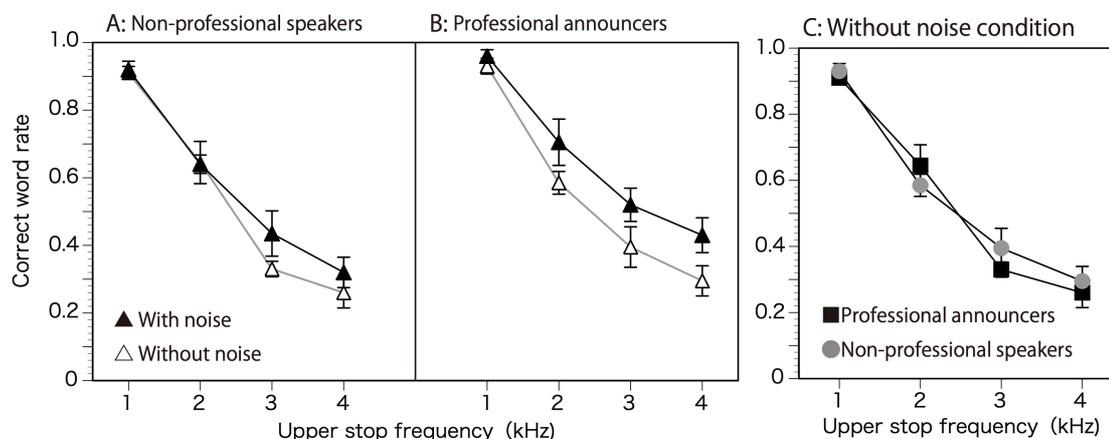


図 2 音韻修復実験の結果 A: 一般人音声のノイズあり条件となし条件の比較. B: アナウンサー音声のノイズあり条件となし条件の比較. C: 一般人音声とアナウンサー音声のノイズなし条件の比較

(3) 聴覚的補完がされやすい音韻と音響特徴

聴取実験の結果から、アナウンサー音声のような雑音下でも聞き取りやすい音声では、多くのスペクトル情報が消失した状況下でも知覚的に音韻が修復されることが示された。特に250～4000 Hzという従来の研究で指摘されたほとんどの音響特徴をカバーできるような帯域が消失しているにもかかわらず、音韻修復がなされていた。そこで音素によって修復のされやすさが異なるのか検証するために、調音法のタイプ別に正答率を算出し、音響特徴が含まれる周波数帯域ごとに整理を行なった。その結果、アナウンサー音声では高周波数帯域タイプと広周波数帯域タイプで高い修復率が認められ、雑音によって知覚的に修復されていることが示された。

どちらの音声も同じ遮断周波数条件でノッチフィルタを重畳していることから、音声の残余成分の量・質に違いがあったことが考えられる。従来の研究からはフィルタ重畳後の残余成分(250 Hz以下、4 kHz以上)には決定的な音響特徴は含まれていないものと考えられる。実際、ノイズなし条件ではアナウンサー音声の了解度は一般人音声と変わらない。しかし、スペクトル情報をみるとアナウンサー音声のスペクトルは高周波数帯域まで広範囲に広がっている。単独では手がかりとはならないが、ノイズが付加されることで手がかり化する情報がアナウンサー音声には含まれていると推測される。

また、従来の研究では音韻の知覚に基本周波数は直接影響することはないと考えられてきた。しかし、本研究で下限遮断周波数を操作し音韻修復への影響を実験的に検討したところ、音韻修復には影響を及ぼす結果が示された。このことから、基本周波数成分は直接音韻の手がかりとはならないが、断片的な音声情報を統合するための手がかりとはなると考えている。これらのことを考慮すると、Glimpsingモデルでは残余成分の周波数帯域によって係数が変わるが、基本周波数帯域の係数を従来よりも重視する必要があることが示唆された。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 小林 まおり, 倉片 憲治	4. 巻 79
2. 論文標題 女声と男声のどちらが聞き取り易いか	5. 発行年 2023年
3. 雑誌名 日本音響学会誌	6. 最初と最後の頁 85-93
掲載論文のDOI (デジタルオブジェクト識別子) 10.20697/jasj.79.2_85	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 小林まおり, 倉片憲治	4. 巻 51
2. 論文標題 音声の聞き取りやすさに関わる性別による音響特徴の違い-ATRオン英データベースを用いた分析-	5. 発行年 2021年
3. 雑誌名 日本音響学会聴覚研究会資料	6. 最初と最後の頁 325-330
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件/うち国際学会 1件）

1. 発表者名 Maori Kobayashi, Masato Akagi
2. 発表標題 Difference of phonemic restoration between professional and non-experts' voices
3. 学会等名 Acoustics 2023 Sydney (国際学会)
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------