

Title	音声生成過程におけるフォルマント変換音声フィードバックの影響に関する研究
Author(s)	齋藤, 和行
Citation	
Issue Date	2004-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1796
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

音声生成過程におけるフォルマント変換音声フィードバック の影響に関する研究

齋藤 和行 (210038)

北陸先端科学技術大学院大学 情報科学研究科

2004 年 2 月 13 日

キーワード: 音声知覚, 発話制御, フォルマント, 変換聴覚フィードバック, リアルタイム処理.

1 はじめに

発話時における聴覚フィードバックの役割については, 音声知覚・生成の相互作用に関する研究として古くから様々な研究がなされてきた. 騒音環境下では, 通常の発話より発話音声が大きくなる現象 (Lombard 効果) や, 発話音声が遅れてフィードバック (遅延聴覚フィードバック) されると, 吃音や発話速度が遅くなる等の現象が生じることは聴覚フィードバックの効果を表す顕著な例である. これらの知見は聴覚フィードバックは発話にとって重要な役割を果たすことを示す証拠として考えられてきた. しかし, これらの報告は定性的な性質を述べるにとどまっており, これらの現象がどのようなメカニズムで生じるかについて説明するには不十分である. また DAF のような発話動作を破壊するような実験では定性的な性質は観察できるが, 発話動作自体が破綻してしまうのでメカニズムを分析するうえで必要な定量的な分析を行うことが難しい. そこで実時間により音響パラメータを変換するような, 定量的な分析を可能にする非破壊的な実験パラダイムの構築が必要となってくる. また発話の全体像を理解するための裏付けられた理論的枠組が必要である. 近年, MRI(核磁気共鳴画像) 等の観測技術発展に伴う神経回路や運動制御に関する多くの発見があったことや信号処理技術や計算機の処理能力の爆発的な向上による実時間の音響パラメータの変換が可能になってきたことから, 音声の知覚・生成の相互作用について調べる上での要件を満たしつつあるといえる.

2 本研究の目的

音声生成・知覚の相互作用に関する問題を運動制御モデルとして扱い, 入力を聴覚刺激, 出力を発話音声とした場合の発話制御モデルを構築するための知見を獲得する. その方法

としてスペクトル構造を変化させたフィードバック音声を与えた場合の応答特性を推定する。スペクトル構造の変化には、被験者の発話音声のフォルマントを変換したものをを用いる。フィードバック音声を呈示し、その時の発話音声を測定することで、発話時の逐次的な応答特性を推定する。

3 実験

実験では発話音声を連続母音 /eaea .../ とし、摂動として発話音声の /e/ の部分を /a/ に変換し、被験者にフィードバックした。発話の開始、及び終了は装着されたヘッドフォンを通じて発話開始信号、発話終了信号により被験者へ通知される。発話開始信号と発話終了信号の間隔は 30 秒間でその間被験者は発話を行う。最初の 10 秒間は”摂動なし”の音声が入力され、次の 10 秒間は変換した”摂動あり”の音声が入力され、最後の 10 秒間は再び”摂動なし”が音声をフィードバックされる。息継ぎは各自のタイミングで自由に行ってよいものとした。これら一連の動作を 1 セットとして各被験者毎に 3 セットずつ行った。各セットの間は休憩として 10 秒間の間隔を設けた。また被験者の意識を聴くことに向け、フィードフォワードによる要因を減少するため、発話音声に変化が現れている間、ボタンを押すというタスクを与えた。

4 結果

分析には発話音声の短時間スペクトルを用いた変動分析を行った。ここではスペクトル推定において、不偏推定量を与える、不偏推定法による分析を行ったその結果 2 名の被験者において摂動を与えた状態で通常時より変動 (分散) が大きくなる傾向が得られた。また実験後の被験者による報告では”摂動なし”のときと比較して”摂動あり”の場合に発話に関する違和感を訴えている。

5 考察

分析の結果からフィードバック音声の変更が、発話動作に何らかの影響を与えていることが推測される。この結果は発話時にフォルマントのようなスペクトル構造に関する情報を利用していることを支持する結果である。そしてそれらの応答は被験者により異なる。その原因として音声習得過程等による個人差が存在するものと思われる。本研究による結論として、通常の発話においては主にフィードフォワードによる制御が行われていると思われる。しかしフォルマントに関する情報を利用している可能性が高いことから、フィードバック制御も同時に行われているものと思われる。つまり本質的には発声の基本周波数制御モデルと同様にフィードフォワード制御とフィードバック制御の並列動作により説明が可能であると推測される。最後に、多分に主観的ではあるが、発話運動の生成における聴覚

フィードバックの役割についての考えを述べる。実験の結果からフィードバックによる影響は発話音声に微弱に現れた程度で発話動作を大きく変化させるものではなかった。これは発話運動を行う上での戦略上、フィードバックゲインを敢えて小さくすることで運動を妨害しないようにして、なおかつフィードバックゲインを完全にゼロにしてしまうのではなく、僅かに残すことにより、環境の変化による外界と外界の内部モデルの誤差を監視し、変化があれば調整するような目的に聴覚フィードバックが用いられているのではないかと考えられる。

6 今後の課題

今回の実験では変換された音声を、補償するような動作を確認するには至らなかったが、被験者は摂動を与えられた状態では発話しにくい等の、発話に対する違和感を訴えていることから、筋電計やカメラによる口の周りの筋肉の動きの測定など音声以外の測定方法を用いることでより明確な応答が確認できる可能性がある。また今回の実験では/ea ea .../という単純な繰り返しの音節であったため、フィードフォワードによる応答がフィードバックによる応答の発生を抑圧したことが考えられる。さらに付け加えると日本語という言語による問題も挙げられる。日本語は母音の数が5つと他の言語に比べて圧倒的に少ない。これはそれだけ母音に関する弁別が他の言語に比べて不利である。これらはフィードバック制御の発生を妨げている要因ではないかと思われる。このため、フィードフォワード制御の要素を除去、もしくは弱めるなどフィードバックの要素を引き出すような実験パラダイムが必要があると考えられる。その方法として普段使わないような、音韻や音節を利用することや、何らかの方法でフォルマントに関する情報を視覚的にもフィードバックしてやることなどが考えられる。また音韻、音節の違い、与える摂動の種類や大きさ、摂動を与えるタイミングなどによる影響についての、定量的な評価は今後の課題である。