

聴覚妨害時の歌唱実験の解析

中村隆志

新潟大学人文学部 〒950-21 新潟市五十嵐2の町 8050

E-Mail: tks@geb.ge.niigata-u.ac.jp

音声、聴覚を用いたコミュニケーションにおいて、自らの発話を聞き取りながらその発話を調整する作用は欠かせない。この自声のフィードバック制御が如何なる性質も持つものなのかを調べるため、被験者の聴覚を妨害するような条件下で、被験者に歌唱を遂行してもらい、聴覚妨害のない条件下での歌唱と比較した。解析の結果、聴覚解放時の歌唱においては、聴覚妨害時に比べ、1：基本周波数成分がより安定していること、2：基本周波数周辺での分散の値がより大きいことが傾向として確認された。このことは聴覚解放時、すなわち自声のフィードバック制御が為されている条件下では、基本周波数が安定している事と同時に基本周波数付近の揺らぎが大きいことを示している。このことは通常の発話時においても、聴取者による補完こそが、会話、コミュニケーションの成立に寄与していることを示している。

キーワード：聴覚妨害、フィードバック制御、ゆらぎ

Analysis of Singing with Auditory Obstruction

Takashi NAKAMURA

Faculty of Humanities, Niigata University

8050 Igarashi-2, Niigata, Japan, 950-21

E-Mail: tks@geb.ge.niigata-u.ac.jp

It is necessary to control vocalization by recognizing one's own speech in verbal and auditory communication. A singing experiment is designed to investigate the feedback feature for the control of one's own voice. Data samples are obtained in the two conditions of singing with open hearing and singing with obstructed hearing. The results of analyses are 1: the spectral envelope around the fundamental frequency (F0) is more stationary in the condition of open hearing, 2: the variance of the peak of F0 is larger in the condition of open hearing than in that of obstructed hearing. Thus, the supplement of a listener may make it possible for dialogue or communication using ordinary vocalization or in the condition with the feedback control of one's own voice.

keywords: auditory obstruction, feedback control, fluctuation

1. 序

音声、聴覚を用いたコミュニケーションにおいて、視覚のそれとの大きな違いは時間の経過の仕方と言えらるだろう[1]。特に発話中の音程や発語の調整という作業は、時間の経過そのものが前提になり、発音する過程においては、事前に如何なる指標をも参照する事ができないままに、発話は遂行され、会話は成立していく。健常者においては、ほぼリアルタイムに近い形で自声のフィードバックを聞く事ができるため[2]、支離滅裂な発話を行うことは意図的な場合をのぞいては、まずあり得ない。一方、先天的障害者においては発声そのものはできても、会話を成立させるに足る発話が可能な人はあまりいない。そのことが、健常者の発話において、リアルタイムのフィードバック制御が行われているかのように思わせる。しかし、フィードバック制御は果たして、それほどうまく為されているのだろうか。今回の実験及びその解析はこの疑問に端を発している。

2. 実験の狙い

一連の対話音声認識の研究によって、会話におけるアクセント、間、語調、韻律その他の持つ情報が明らかにされつつある。それらが示すことの一つは、発話の遂行がその聴取者の補完があって初めて成立することである。聴取者はその聴覚をもって対話者の発話行為に参画することになる。この聴取者側の補完が発話制御の不全を覆っていることも考えられよう。発話者自身も聴取者でもあるため、そのことはなかなか表面化しないだろう。

そこで、上述の疑問の形を変えてみる。歩行するに支障のない健常者を考えてみよう。彼は視覚的にも健常者であるとする。多くの読者には我が事と考えてもらって差し支えない。広場に幅10センチ(20センチでも良い)の

直線が延々と続いており、彼にその直線上を、左右の足がそれぞれ半分以上直線からはみ出ないように歩くように命じたとする。彼はおそらく言われたとおりの歩行が可能だろう。歩行に関する健常者の身体は、メカニカルに上の命令の遂行が可能である。視覚的な情報からの制御も遂行されることになる。彼は何十メートルでも歩行可能だろう。そこで、状況を少し変えてみる。かの直線上に幅10センチ(または20センチ)の平均台をおき、その上で上述の命令を遂行することを命じたとしよう。広場の直線上の歩行と本質的な違いはない。歩行のメカニズムも視覚情報による制御もそのまま行えばよいのである。しかし、何らかの訓練を受けていればともかく、彼はすぐに落下するだろう。

上のは単なる寓話ではない。制御のための明確な指標が存在したとしても、制御が不完全であることの例である。広場の直線と平均台の対比は心理的要因などを問うのではなく、人体のメカニカルな可能性と視覚による制御の可能性を示し、かつその実現が不全となることを引き合いに出すためである。彼が落下するのは体のずれを補正しようとして補正しきれない、または補正しすぎてしまう「読み違い」が起こるからである。この読み違いは以前のずれに起因しながら、それに付加する形で新たなずれを引き起こす。このずれの補正においても読み違いはさらに起こる可能性を常に持っている。読み違いは事後になって初めて顕現する。よってこの読み違いとずれ生成のサイクルは不断におこり続ける可能性を持っている[3,4]。

翻って、発話におけるフィードバック調整においても同様な読み違いとずれ生成が起きている可能性がある。このずれがおこる要因はまさに自声のフィードバックを聞いていることである。もし、ずれ生成が自声のフィードバックによって起こっているとすれば、そのことを確かめるためには、そのフィードバックを遮断すればよい。上述の平均台の寓話を用いるなら、平均台上の歩行者が目隠して歩く図

を想像してもらえばよい。彼は何とか落ちまいとバランスをとろうとして、その結果として落下するというのではなく、平然と歩き続けるまま、足を踏み外す形で落下することだろう。

この読み違えとずれ生成を明示的にするためには、実験方法として、ずれないように発話する事が求められる行為を探し、それを被験者にやらしてもらえばよい。そこで、歌唱を実験材料として選択した。また、自声のフィードバックの遮断は骨伝導があるため、完全には行かないが、ある程度、攪乱する方法があればよい。このことを実現すべく、以下の実験法を考案した。

3. 実験法

本実験を簡略に言うならば、異なる条件で被験者にカラオケ歌唱を歌ってもらい、それをサンプリングして解析した、となる。異なる条件とは、

条件0：そのまま、自声のフィードバック有り
で歌唱、

条件1：大音響のヘッドホンを付けることで不完全ながらも自声のフィードバックを遮断したままの歌唱、

の2つの条件を示す(図1)。

図1において、被験者のヘッドホンは3つの役割を果たす。第1にカラオケ伴奏を聞かせることにより、歌うべき音程を教示する。第2にリズムを教示する。自声のフィードバックの遮断によるリズム感のずれの発生も十分検討すべき課題ではあるが、今回は音程の調整のずれを計測するため、リズムのずれをなるべくなくすような環境にした。第3に大音響による自声のフィードバックの妨害である。これによって、被験者が許容できる範囲の大きさで自分の声がよく聞こえない状況を作った。さらに、被験者の歌唱のサンプリング時にカラオケ伴奏まで混入してはデータにならないため、ヘッドホンによる伴奏がもっとも妥当な方法と考え

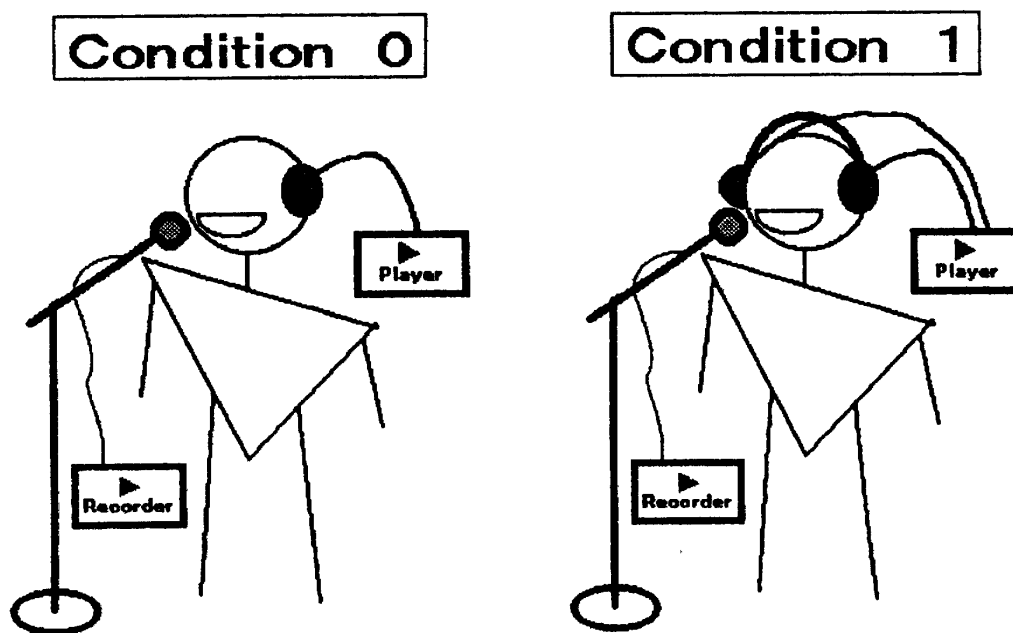


図 1. Experiment Set-up

られる。条件0においてもカラオケ伴奏を聞かせる状況を作るため、ヘッドホンをさせて、片側のホーンを耳からわずかにずらす形にした。これにより、片耳からカラオケ伴奏、もう片耳から自声のフィードバックの両方を確保する方法を採った。

サンプリングについて

1 : サンプリングレート : 44,100[Hz]

2 : サンプリング語長 : 8 bit

3 : モノラル録音

の条件下で行った。

被験者は12人で、ひとりには高校の音楽(ピアノ)の教員であり、残り11人は高校生で器楽部の部員である。歌唱の専門家はいない。使用した曲は

1 : 四季の歌

2 : 乾杯

3 : 贈る言葉

4 : てんとう虫のサンバ

の4曲で、曲に合わせて、1と4のデータを2等分、2と3のデータを3等分して約30秒弱の大きさにしたものをデータとして用いた。データ数は12人分で120個になる。

4. 解析結果

各データについて2048点ずつFFTを行った。FFTの結果を基に基本周波数周辺のピークの大きさ、及びピークの分散の大き

さを求め、120個のデータについて各データ毎に時間を通して積算し、その結果を条件0と条件1について比べてみた。その結果を表1に示す。また、3倍音から8倍音までのピークの周辺についても同様にピークの大きさ及び分散の大きさを求め、120個のデータ毎について積算し、条件0と条件1について比較した。その結果を表2に示す。

分散を求める際のガウス近似の粗さを考慮して、FFTスペクトルのピークの形を定量化するもう一つの指標Peak Thickness(PT)を定義した。

$$PT(f) = (1/A(f)) \sum_{i=-K}^K i^2 A(f+i),$$

s. t. f : frequency,

$A(f)$: Spectral Envelope.

これを基本周波数及び倍音周辺におけるスペクトルの最大値をとる周波数に適用する。最大値そのものは i の項の積により、積算値には無効である。ここでは、 $K=3$ として計量した。分散値と同様、基本周波数周辺及び3倍音から8倍音のピークに対して、PTを適用し、120個のデータ毎に積算した。結果は分散値の場合とほぼ同じ傾向が出た。結果を表3に記す。

	条件0>条件1となる割合
ピークの大きさ	111 / 120 = 92.5 %
ピークの分散	110 / 120 = 91.7 %

表1. 基本周波数周辺のピークの比較

	条件0>条件1となる割合
ピークの大きさ	8 / 120 = 6.7 %
ピークの分散	6 / 120 = 5 %

表2. 3倍音から8倍音のピークの比較

	条件0>条件1となる割合
基本周波数周辺	111 / 120 = 92.5 %
3倍音から8倍音	7 / 120 = 5.8 %

表3. 各ピークのPT値の比較

以上をまとめると次のようになる。基本周波数周辺のピークの大きさについては条件0の方が大きくなる傾向がある。これは基本周波数の音が高音に割れたり、声がひっくり返ったりすることなく、歌唱がより安定していることを示す。その一方で分散の値も条件0の方が大きい傾向がある。これはただ、歌唱が安定しているだけでなく、基本周波数がより揺らぎ続けていることを示す。また、3倍音から8倍音ぐらいまでの周波数域においては条件0と条件1の関係は逆転する。これは、条件1の方がピークが大きくなり、分散もより大きくなる傾向がある。この結果は先の表1の結果と総合的であり、条件1での歌唱が不安定であることを裏付ける。

5 : むすび

音声、聴覚を用いたコミュニケーションにおいては、自声のフィードバックを聞く事による発話制御がなされており、それと並行して会話が進行していく。このフィードバック制御が如何なる特徴を持つのかを調べるため、敢えて自声のフィードバックを妨害してやるような状況を作ってやり、妨害のない状態とあわせて歌唱実験を行った。結果、フィードバックの妨害のない状態、つまり健常者の通常時の歌唱の方が、妨害のある状態に比べて、基本周波数

が安定しており、高周波数域への移行が起こりにくいことがわかった。また、同時に基本周波数のピーク周りの分散を調べてみるとフィードバック妨害のない状態の方が分散の値が大きかった。このことは、音程の安定とその不断の揺らぎが、自声のフィードバック制御によって、まさに同時に得られるものであることを示す。今回の実験の結果が対話音声認識、コミュニケーションの解析に何らかの示唆を与えることが期待される。

謝辞：普段、議論いただく神戸大学理学部、郡司幸夫先生に感謝します。また、データ採取を快く引き受けて下さった大阪市立西淀川高校、蔭木恵教諭及びその教え子達に感謝します。

引用文献

- [1]視聴覚情報科学、ATR基礎技術研究所編、オーム社、1994
- [2]音声・聴覚と神経回路網モデル、甘利俊一監修、オーム社、1990
- [3]プロトバイオロジー、松野孝一郎訳著、東京図書、1989
- [4]生命と時間、そして原生-計算と存在論的観測、郡司ベギオ幸夫、現代思想、vol. 24-13, 1996