

Title	歌唱における気導・骨導の聴覚フィードバックに関する調査研究 [課題研究報告書]
Author(s)	鳥谷, 輝樹
Citation	
Issue Date	2016-09
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/13747">http://hdl.handle.net/10119/13747</a>
Rights	
Description	Supervisor: 鶴木 祐史, 情報科学研究科, 修士

課題研究報告書

歌唱における気導・骨導の聴覚フィードバック  
に関する調査研究

北陸先端科学技術大学院大学  
情報科学研究科情報科学専攻

鳥谷 輝樹

2016年9月

## 課題研究報告書

# 歌唱における気導・骨導の聴覚フィードバック に関する調査研究

指導教員 鵜木祐史 准教授

審査委員主査 鵜木祐史 准教授  
審査委員 赤木正人 教授  
審査委員 党建武 教授

北陸先端科学技術大学院大学  
情報科学研究科情報科学専攻

1410031 鳥谷 輝樹

提出年月: 2016年8月

## 概要

ヒトの音声コミュニケーションの一つとして歌唱があり，世界中で親しまれている．しかしながら，日常会話に問題がないにも関わらず，上手に歌唱することができない「調子外れ」の歌唱者もいる．こうした歌唱者が上手に歌唱できるようになるためには，ヒトの歌唱メカニズムの理解が不可欠である．これまで，歌唱時の音声知覚・音声生成に関する研究が行なわれており，知覚と生成の相互作用の観点にも注目されている．

知覚と生成を相互に関係づける機構として，聴覚フィードバックがある．聴覚フィードバックが発話や歌唱に与える影響について，気導音声に着目した検討が行われている一方で，骨導音声としての聴覚フィードバックによる発話・歌唱への影響については考慮されていない．骨導の聴覚フィードバックが上手な歌声の創出に寄与しているのかどうかは，未だ明らかにされていない．

本研究では，歌唱における知覚・生成プロセスの解明を目指すために，気導・骨導それぞれの聴覚フィードバックが歌唱に与える影響に着目する．本調査研究の目的は，気導・骨導の聴覚フィードバックがそれぞれ歌唱とどのように関係しているかについて，関連研究の知見をもとに予想を立てることである．そのため本論文では，歌声に見られる音響的特徴の調査，発話や歌唱における気導の聴覚フィードバックに関する調査と，骨導音声の知覚や音響的性質に関する調査を行ない，気導・骨導の聴覚フィードバックと歌唱との関係を考察する．

まず，歌声の音響的特徴に関する研究を調査した．調査結果から，歌声は話声よりも長い音韻持続時間，強い音声パワーが見られ，歌声特有の音響的特徴としてF0軌跡とF0動的変動（ヴィブラート等），Singer's Formantがあることが分かった．調子外れに関連する特徴として，F0軌跡とF0動的変動が特に重要であることが分かった．

次に，聴覚フィードバックに関連する現象と手法，および，歌声を対象とした聴覚フィードバックに関する関連研究の調査を行なった．最初に，周囲の騒音レベル変化に応じた発声強度・F0の変化（ロンバード効果）があることが分かった．また，聴覚フィードバックの知見は話者の生成音声に遅延や変形を施して話者に知覚させる手法により得られ，遅延聴覚フィードバック（DAF）による発話の乱れや変形聴覚フィードバック（TAF）に対する補償・追従反応が見られることが分かった．歌唱における気導の聴覚フィードバックに関する関連研究は，聴覚フィードバックの変形，および，話者内部の手がかりへの着目という二つのアプローチがあることが分かった．前者のアプローチの検討から，生成音声のF0とターゲット信号との間でF0の逸脱を補償することが分かった．また，ターゲット信号へのF0追唱の精度は，歌唱熟達者の方が高いことも分かった．さらに，F0の周期的変動を付与した刺激に対応して，歌声のF0にも変動が見られることが分かった．後者のアプローチの検討として，聴覚フィードバックを遮断するために歌唱時に気導から雑音を呈示すると，意図したF0からの逸脱が増加する傾向と，逸脱が減少するという2つの傾向が見られ，歌唱時には話者内部を伝わる感覚情報がある可能性があることが分かった．

上記の知見はすべて、気導の聴覚フィードバックを前提として得られたものであり、骨導の聴覚フィードバックが歌唱時の F0 やその動的変動に果たす役割については依然として未解明であることが分かった。

最後に、気導音声に加え骨導音声に着目するため、骨導音声の知覚と音響的特徴に関する研究の調査を行なった。骨導音声は主に 5 つの知覚経路を持つことが分かった。音響分析から、骨導音声には F0 付近の成分は十分に含まれる一方、4 kHz 以上の高域成分は低下している可能性があることが分かった。生理的・心理物理的検討から、骨導音声は 0.7 ~ 2 kHz 付近の周波数域の知覚に主に寄与していることが分かった。関連研究から、喉頭振動・聴覚の両方によるフィードバックは聴覚フィードバックのみの場合よりも強く、話者の F0 制御に影響していることが分かった。また、喉頭振動刺激による F0 の補償反応は聴覚刺激によるものよりも低い潜時を持つことも分かった。

結論として、これまでの知見を踏まえ、歌唱時の F0 とその動的変動に対して気導・骨導の両方の聴覚フィードバックが影響し、それぞれの優位性は歌唱者により異なるという予想を立てた。予備検討として、気導・骨導の DAF の下での発話スタイルを調査し、骨導呈示時での発話スタイルの変化が特に顕著であることを示した。今後は予想を実証するため、骨導音声に関するデバイスの特性調査と、気導・骨導音声の知覚の検討、気導・骨導音声の撮動呈示が歌声に与える影響について検討する予定である。

# 目次

第1章	序論	1
1.1	はじめに	1
1.2	聴覚フィードバックと気導・骨導	2
1.3	本調査研究の目的	4
1.4	本報告書の構成	4
第2章	歌声の音響的特徴	5
2.1	音声に見られる音響的特徴	5
2.1.1	時間的特徴	5
2.1.2	F0とその変動	5
2.1.3	スペクトル特徴	6
2.2	歌声特有の音響的特徴	6
2.2.1	音韻持続時間と音声パワー	6
2.2.2	F0軌跡・F0の動的変動	9
2.2.3	Singer's Formant	11
2.3	歌声の音響的特徴のまとめ	13
第3章	聴覚フィードバックの関連研究	14
3.1	聴覚フィードバックに関連する現象と手法	14
3.1.1	ロンバード効果	14
3.1.2	遅延聴覚フィードバック (DAF)	16
3.1.3	変形聴覚フィードバック (TAF)	16
3.2	歌声を対象とした関連研究	19
3.2.1	聴覚フィードバックの変形	19
3.2.2	話者内部の手がかりへの着目	23
3.3	これまでの知見のまとめ	25
第4章	気導・骨導の聴覚フィードバック	27
4.1	骨導音声の音響的特徴と知覚特性	27
4.2	気導・骨導による歌声の特徴伝達に関する考察	31
4.3	気導・骨導の聴覚フィードバックの性質を明らかにするための指針	33



# 目次

1.1	ことばの鎖 (Speech Chain) . . . . .	3
1.2	聴覚フィードバック (気導・骨導) . . . . .	3
2.1	発声強度の時間変化による音声強調パターン . . . . .	7
2.2	母音発声時のスペクトログラム . . . . .	7
2.3	母音における第一・第二フォルマントの関係 . . . . .	8
2.4	話声と歌声の F0 軌跡 . . . . .	10
2.5	歌声の F0 動的変動 . . . . .	10
2.6	話声・歌声のスペクトル包絡の概形 . . . . .	12
3.1	話者内部における音声知覚と音声生成の関係 . . . . .	15
3.2	変形聴覚フィードバック (TAF) の実験方法の概略 . . . . .	18
3.3	F0 の上昇刺激に対する補償反応の例 . . . . .	18
3.4	追唱課題に対する歌声の F0 推移 . . . . .	20
3.5	F0 の変調刺激と生成音声の F0 軌跡との関係の概略 . . . . .	22
3.6	雑音呈示/非呈示時における F0 逸脱の比較 . . . . .	24
3.7	EF/IF 条件における F0 逸脱の比較 . . . . .	24
4.1	骨導音声の伝達経路の概略 . . . . .	28
4.2	気導・骨導音声の周波数特性の比較 . . . . .	30
4.3	気導・骨導・喉頭振動のフィードバックによる F0 制御の概略 . . . . .	32
5.1	DAF の実験機構の概略 . . . . .	40
5.2	音声波形中の $T_i$ , $T_{d,i}$ ( $d = 200$ ) および $s_{d,j}$ の例 . . . . .	41
5.3	遅延条件による $R_d$ の推移 . . . . .	42
5.4	遅延条件による $N_d$ の推移 . . . . .	43

# 表 目 次

2.1	話声と歌声の音響的特徴の違い . . . . .	13
4.1	骨導音声の知覚経路と寄与する周波数成分 . . . . .	28

# 第1章 序論

## 1.1 はじめに

ヒトの音声コミュニケーションには、日常会話や歌唱がある。歌唱には、言語的な音声に抑揚がつけられたことから始まったといわれている「言語起源説」だけでなく、言語とは無関係に感情を表出したことから始まったといわれている「感情起源説」がある [1]、また、一説では音楽の起源が歌唱であると考えられている [2]。歌唱は発声・調音器官を用いた音声コミュニケーションという意味で日常会話と共通しているが、上記のように、元来それら二つは別々の成り立ちや用途を持っていると考えられている。さらにヒトは、歌声を話声とは別のものとして知覚・生成できる。歌唱や歌声への関心は今日非常に高まっており、その代表としてYAMAHA(株)のVOCALOID [3]などの歌声合成システムや、高機能な自動カラオケシステムがある。

ヒトが歌唱するとき、如何に歌を上手に歌えるようになるか、ということは重要な関心事の一つである。このことに基づき、ヒト自身が歌唱を行なう際のメカニズムの解明を目指して、知覚的側面（心理学）や生成的側面（生理学）などからの検討が行なわれている。

前述のように、歌唱時には日常会話時と同様に発声・調音のプロセスを経て音声生成が行なわれる。一方で、歌唱者によっては、日常会話には問題がないにも関わらず、上手に歌唱を行うことができない場合がある。特に、歌唱時の音程やリズムを適切に保つことができない状態は調子外れ（俗に「音痴」）と呼ばれる [4,5]。実際に、調子外れは日常会話時とは独立して、歌唱時に特有の問題である可能性が指摘されている [6]。これまで、調子外れを引き起こす要因は、音声知覚の不全である感覚性のもの（心理的要因） [7,8] と、音声生成の不全である運動性のもの（生理的要因） [9] として区別されると考えられてきた。近年では、音声知覚・音声生成の個々に問題がなくても、知覚・生成の相互作用の中で調子外れを引き起こす要因があるのではないかと考えられている [10-12]。このことに基づき、調子外れの改善および歌唱能力の向上を狙いとして、歌唱時の知覚・生成の相互関係に着目してそれらの関係を視覚的に呈示する試みが提案されている [13-15]。

ヒトが調子外れを起こさずに上手に歌唱できることを目指すためには、まず音声を知覚し、次に意図した音声を適切に生成することが必要となる。さらにそれだけでなく、知覚と生成の間で適切な相互作用を持つということも、歌唱時に必要不可欠となる。したがって、歌唱中の知覚・生成間の相互関係に着目し、それらの間の処理過程を明らかにしていくことが重要であると考えられる。

## 1.2 聴覚フィードバックと気導・骨導

歌唱に限らず，ヒトの音声コミュニケーションにおける話者間・話者内での知覚・生成間の情報交換の通路は，ことばの鎖（Speech Chain）と呼ばれる [16]．ことばの鎖の概念図を，図 1.1 に示す．ことばの鎖でやりとりされる情報は，音響レベル・生理レベル・言語レベルの 3 つに分けられる [16]．ことばの鎖による情報交換の間，各レベルにおいて発話のモニタリングを行なうループが存在すると考えられている [17]．この中で，音響レベルでのモニタリングループは聴覚フィードバック（Auditory Feedback）と呼ばれ，音声の知覚と生成を密接に関係づけるものである．すなわち，話者は他者の音声を知覚して話者自身の音声を生成するだけではなく，話者自身が生成した音声を話者自身が知覚している．この仕組みにより，話者は音声の物理的性質が意図通りになるように生成系を制御することができる．その証拠として，ライブ会場やカラオケ等では歌唱者自身に歌声を呈示することにより安定した歌唱を実現している．

聴覚フィードバックにより話者が知覚する自身の音声には，体外で空気を介して外耳道に伝わる気導音声と，体内で骨振動により直接聴覚系に伝わる骨導音声がある．気導・骨導による聴覚フィードバックの概念図を図 1.2 に示す．これまでの関連研究から，聴覚フィードバックに関する現象やさまざまな性質が明らかにされている [18]．しかし，これまでの研究では気導の聴覚フィードバックのみに着目しており，骨導の聴覚フィードバックが音声や歌声に与える影響は考慮されていない．もしも，骨導の聴覚フィードバックが歌唱に重要な役割を果たしているのであれば，気導音声の知覚を前提としたこれまでの知見からは，歌唱中の知覚・生成間の相互作用を十分に明らかにすることはできないと考えられる．

実際に，歌唱者自身の骨導音声は，かなりの大きさと知覚されていると述べられている [20]．また，声楽家の間では，歌唱者自身が歌声を知覚（Self-Listening）する過程において，骨導の聴覚フィードバックが重要な役割を果たしていると主張されている [19]．これらのことを踏まえると，歌唱における知覚・生成の相互関係をより深く明らかにするためには，気導・骨導の聴覚フィードバックのそれぞれの性質について検討すべきである．気導・骨導それぞれの聴覚フィードバックが歌声に与える影響を解明することができれば，歌唱の場面に限らず，ヒトの発話における知覚・生成メカニズムに関する知見をさらに深化することができる．また，発話障害者への治療・支援などへの知見の応用も期待される．

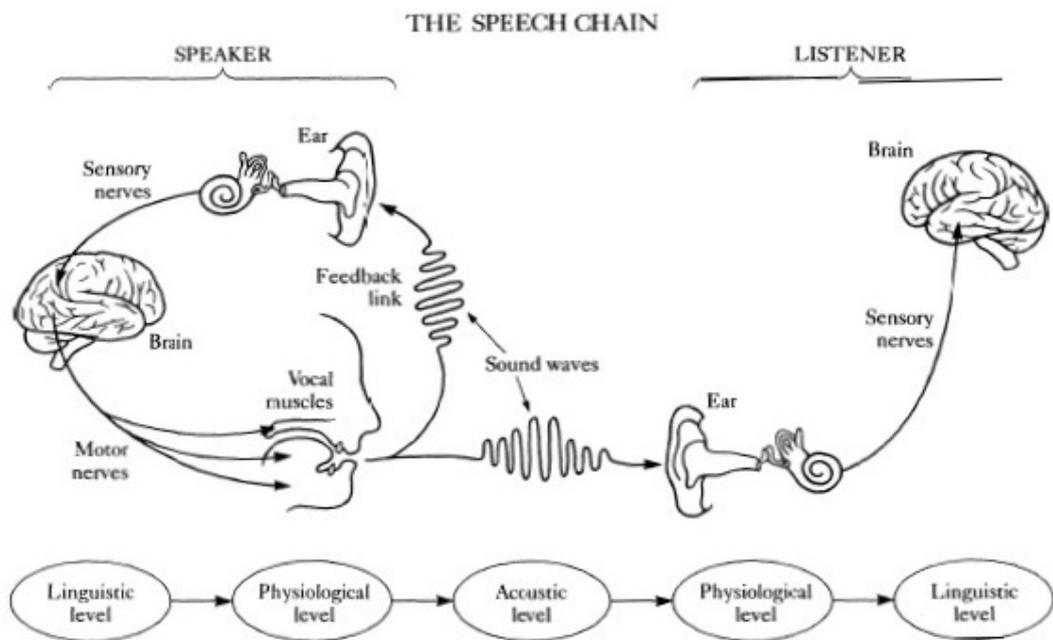


図 1.1: ことばの鎖 (Speech Chain, 文献 [16] より引用)

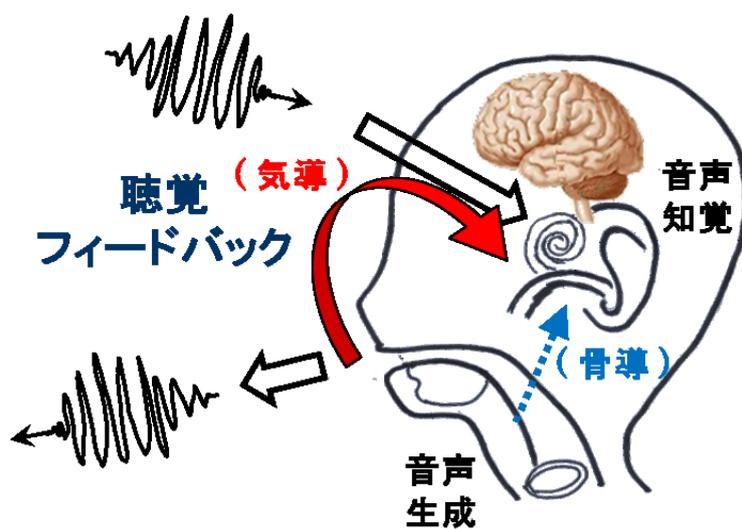


図 1.2: 聴覚フィードバック (気導・骨導)

### 1.3 本調査研究の目的

研究の最終的な狙いは、気導・骨導の聴覚フィードバックのそれぞれに着目することで、歌唱における知覚・生成プロセスを解明することにある。本調査研究の目的は、気導・骨導それぞれの聴覚フィードバックが歌声に与える影響について、関連研究の知見をもとに予想を立てることである。そのため本論文では、まず、歌声の音響的特徴に関する関連研究を調査する。次に、聴覚フィードバックが音声・歌声に与える影響について、気導音声に着目した関連研究を調査する、そして、骨導の聴覚フィードバックに着目するために、骨導音声の知覚・音響的性質に関する研究を調査する。最後に、調査結果をもとに、気導・骨導それぞれの聴覚フィードバックが歌声に与える影響について考察を図る。

### 1.4 本報告書の構成

本報告書は5章で構成されている。第1章では、本研究における問題点を明確にするため、研究背景と研究の着眼点、本研究の狙いと本調査研究の目的を述べ、序論とする。第2章では、歌唱時の聴覚フィードバックを議論する上で必要となる、音声・歌声の音響特徴について述べる。第3章では、聴覚フィードバックが音声・歌声に与える影響について、気導音声に着目された関連研究について述べ、明らかになったことと未解明な点について整理する。第4章では、骨導音声の知覚・音響的特徴に関する研究について述べる。そして、調査により得られた知見を踏まえて歌唱における気導・骨導のフィードバックの性質について考察する。第5章では、本報告書の調査結果と、予想される気導・骨導の聴覚フィードバックの性質を述べ、結論とする。

## 第2章 歌声の音響的特徴

### 2.1 音声に見られる音響的特徴

音声生成系の機能は、声帯による音源生成と声道による調音、口唇からの放射の組み合わせで表現できる [21]。この考え方に基づく音声生成の機能モデルは、音源フィルタモデル (Source-Filter Model) と呼ばれる [22]。このモデルは声帯振動を音源とし、時々刻々変化する声道の共鳴特性をフィルタの機能と見なした音声生成モデルである。声帯振動に由来するパラメータとして基本周波数 ( $F_0$ ) および喉頭スペクトルがあり、声道形状に由来する共鳴特性、口唇から放出される際の伝達特性 (放射特性) によってスペクトル形状が時間変化する [23]。本節では上記のような音声生成プロセスを踏まえ、音声の時間的特徴に加えて  $F_0$ ・スペクトルのパラメータに関する特徴を述べる。

#### 2.1.1 時間的特徴

発声強度の時間変化に対応した音声の強調パターンは言語情報を表す重要な特徴であり [23]、例として図 2.1 がある。図 2.1 は “import” (名詞) と “import” (動詞) を続けて発音した際の音声波形 (1 段目)、整流波形 (2 段目) と、20 ms および 80 ms の窓長に対応した RMS 振幅曲線 (3, 4 段目) を表している。言語情報の異なる 2 単語は、発声強度に関して互いに異なる時間変化を有することが分かる。時間に関連するその他の音響的性質として、音韻の持続時間がある。Port & Dalby は、英語の音声において子音/母音の時間比に対応した音声強調パターンが見られることを示している [25]。

#### 2.1.2 $F_0$ とその変動

音声の  $F_0$  は、声帯振動に由来する音響パラメータである。言語・非言語情報を表現する上で、音声の  $F_0$  とその時間的変動 ( $F_0$  軌跡) も重要な音響的特徴である。Ng らは、通常の喉頭発声者、人工喉頭発声者および食道発声者の音声をを用いた知覚実験を行ない、言語情報伝達に寄与する音響的特徴を調査した。この結果から、発声強度と音韻持続時間よりも、 $F_0$  軌跡が言語情報理解に対する音響的手がかりとして支配的である可能性が示唆されている [26]。また、感情等の非言語情報の違いによっても、 $F_0$  軌跡のパターンが大きく変化することが分かっている [27]。

### 2.1.3 スペクトル特徴

声道の共振特性に由来する，スペクトル上のピーク（フォルマント）は，特に音声の音韻情報伝達に大きく寄与している．図 2.2 は，3 種類の母音発声時のスペクトログラムを示しており，縞模様の濃い部分は周波数の低い順に第一フォルマント（F1），第二フォルマント（F2），と対応している．音韻情報の違いに応じて F1 と F2 の値およびその軌跡が変化していることが示されている [28]．話者が発声した母音と，そのときの F1，F2 との関係性を F1-F2 平面に表すと，図 2.3 のようになることが知られている [29]．この関係から，音声のスペクトル特徴が音韻情報と対応していることが分かる．

## 2.2 歌声特有の音響的特徴

歌声は音声の一形態であるという面で，前節で述べた音声の音響的性質を含有している．一方で，ヒトは歌声を話声とは異なるものとして知覚することができる．歌声を評価するに当たり，リズムや音の高さ（ピッチ），音の大きさ（ラウドネス）等の知覚が重要視されている [30]．歌声を知覚する際の心理的印象に対して，歌声のどのような物理的性質が歌声を特徴づけているのかを明らかにすることを狙いにして，1970 年代ごろより歌声の音響分析が盛んに行なわれるようになった [31,32]．これまで，一般の音声と対比させる形で，歌唱時の音声に特化した音響的性質が調査・報告されている．本節では，歌声に特有の音響的性質について，時間的特徴および F0・スペクトルのパラメータに着目して整理する．

### 2.2.1 音韻持続時間と音声パワー

Livingstone らは話声・歌声という異なる様式における音響的類似性，相違性に着目し，様式（話声・歌声）と複数の感情を変化させた俳優の収録音声进行分析した．この検討では，種々の音響的特徴の違いに関して様式・感情による 2 要因分散分析を行なった結果，話声・歌声の様式の違いにより有意に異なる音響的特徴として音韻持続時間および平均音声強度があることを示した [33]．歌声を特徴づける音響的性質に，話声よりも長い音韻持続時間と，強い音声強度が挙げられると考えられる．Nakano & Goto による歌声から歌声への音声合成システム「VocaListener」[34]でも音韻持続時間と音声パワーに着目している．このシステムでは，対象とする歌声の分析時に音韻情報と歌詞情報を同期させることで各音韻の開始点と持続時間を推定している．ここではさらに，歌声の音声パワーも推定され，分析-合成間でこれらの推定処理を繰り返すことで歌声の再合成を実現している．

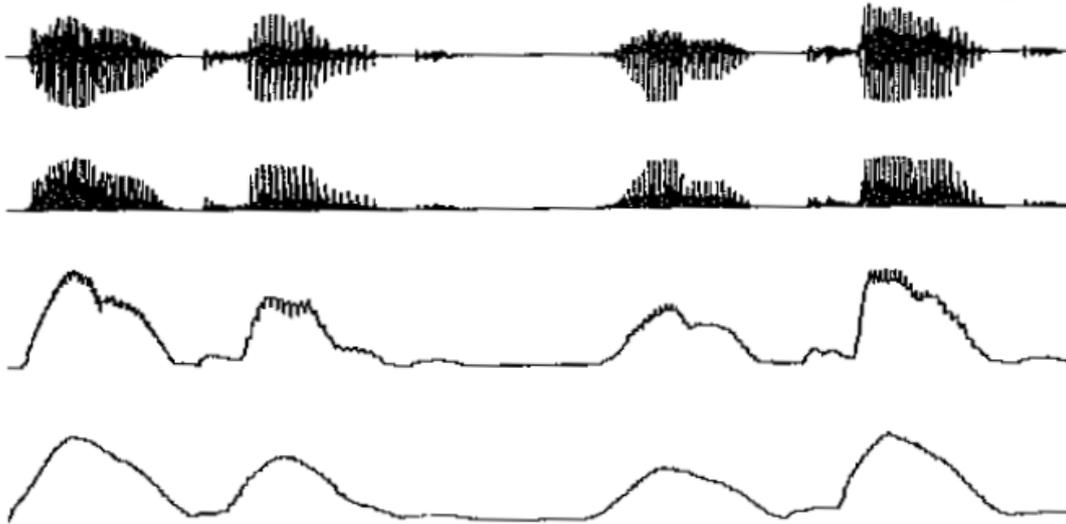


図 2.1: 発声強度の時間変化による音声強調パターン：音声波形（1 段目），整流波形（2 段目），RMS 振幅曲線（3, 4 段目）（文献 [23] より引用）

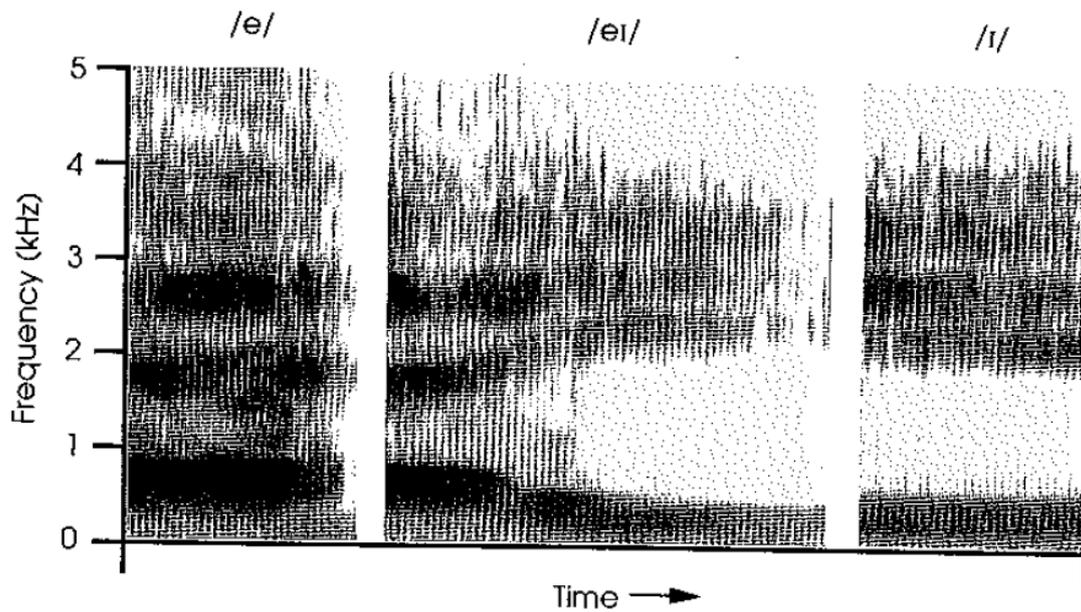


図 2.2: 母音発声時のスペクトログラム（文献 [28] より引用）

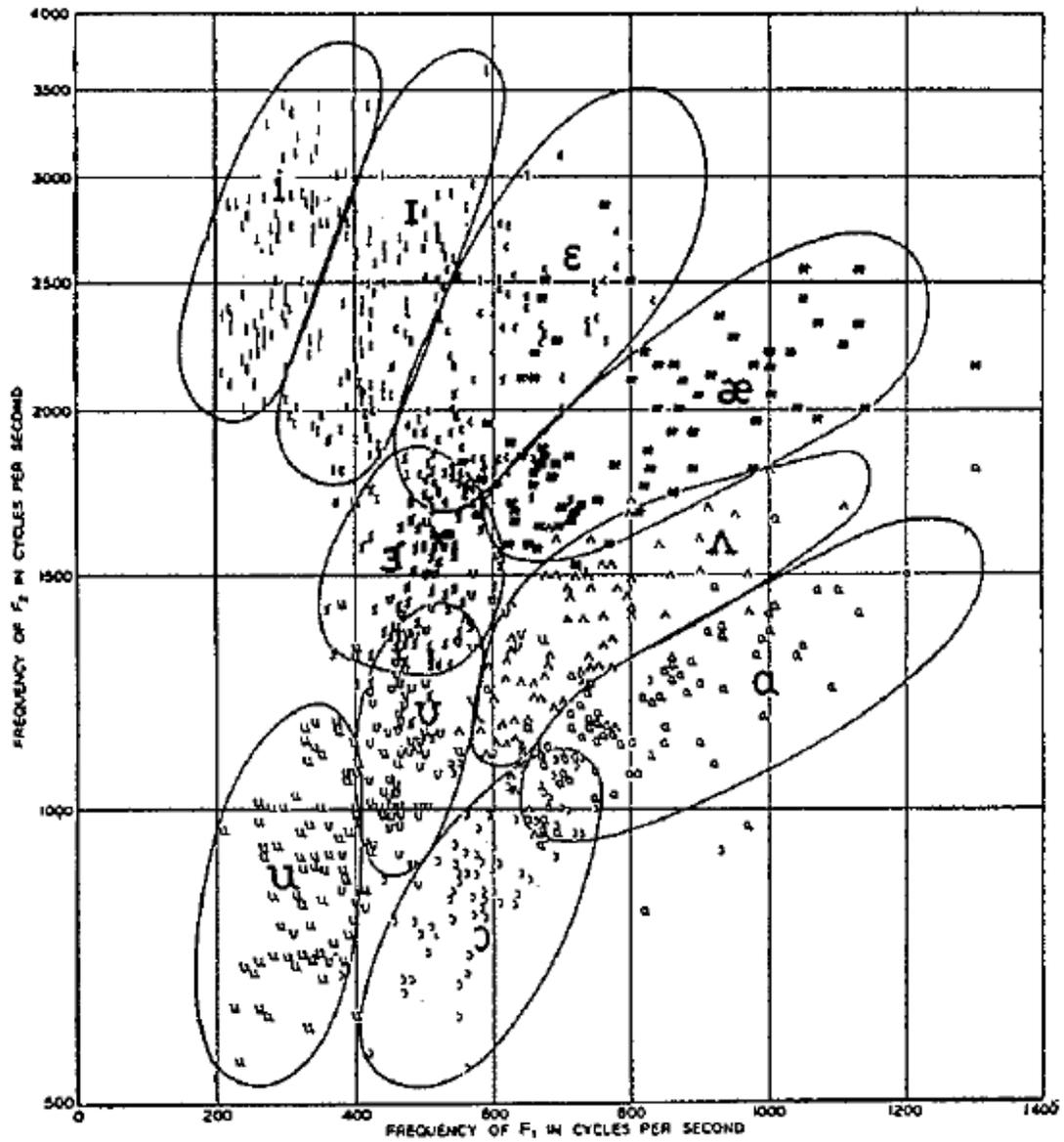


図 2.3: 母音における第一・第二フォルマントの関係 (文献 [29] より引用)

## 2.2.2 F0 軌跡・F0 の動的変動

大石らは、朗読音声と歌声の識別に寄与する音響的特徴の検討を狙いとし、信号長を変化させた際の朗読音声と歌声について、被験者による手動識別および複数の手法による自動識別を行なった。結果として、被験者による手動識別の精度が最も高い傾向にあったが、特徴量として F0 軌跡・スペクトル包絡の時間変化を利用した自動識別手法によっても高い識別精度が得られたと報告されている [35]。この結果から、F0 の時間変化も、歌声を特徴づける音響的性質の 1 つであると考えられる。実際に Zatorre & Baum により話声・歌声の F0 軌跡が分析された。その分析結果を図 2.4 に示す。図 2.4 から、話声では F0 が滑らかに推移しているのに対して、歌声では F0 が階段状に推移するような軌跡を描いていることが確認できる [36]。歌唱では、ある定まった音階を保つように発声する必要があり、話声と歌声の発声スタイルの違いは主に F0 軌跡の違いとして表れていると言える。剣持は、話声と歌声を明確に区別する要因として、音程とタイミングが楽譜に支配されることを指摘している [37]。これは、音程を規定する物理量である F0 が歌声の重要な要素の 1 つであることを意味する。歌声合成時にも F0 の情報は特に重要視されており、先に挙げた VocaListener [34] においても持続時間・音声パワーとともに F0 が着目され、分析処理時に歌声の F0 推定が行なわれて楽音内の適切なピッチ情報に変換される。また、楽曲からの歌声情報の分離抽出に着目した研究では、歌声の F0 を推定することで歌声の調波成分を取り出し、歌声を再合成している [38]。

歌声における F0 の高低の範囲に応じて、いくつかの声質を持つと考えられている。その域内で同様の性質を持つと知覚される発声周波数域は、声区と呼ばれている [20]。Sundberg は、発声時の声区の境界は一般に男性で 200 ~ 350 Hz 付近に、女性で 400 Hz 付近と 660 Hz 付近にあることを述べている [20]。また、声区の数には性差や個人差があることも知られている [20]。

上記の特徴は F0 の長期的な変化に着目したものであるが、歌声の F0 には動的な周期的変動が見られることも知られており、これはヴィブラートと呼ばれる [39, 40]。歌声の心理的印象と音響的特徴との関係に着目した齋藤らの検討から、ヴィブラートとそれに同期する振幅包絡変動が歌声の「揺れ」の印象と対応していることが明らかとなっている [41]。Scherer らは話声・歌声の感情表現の違いに着目し、主観評価の結果から歌声創出時にヴィブラート等の微小摂動の使用頻度が高いことを示している [42]。またその他に、図 2.5 のようなオーバーシュート、プレパレーションと呼ばれる F0 の動的変動も歌声特有の音響的性質として見られることが分かっている [43]。

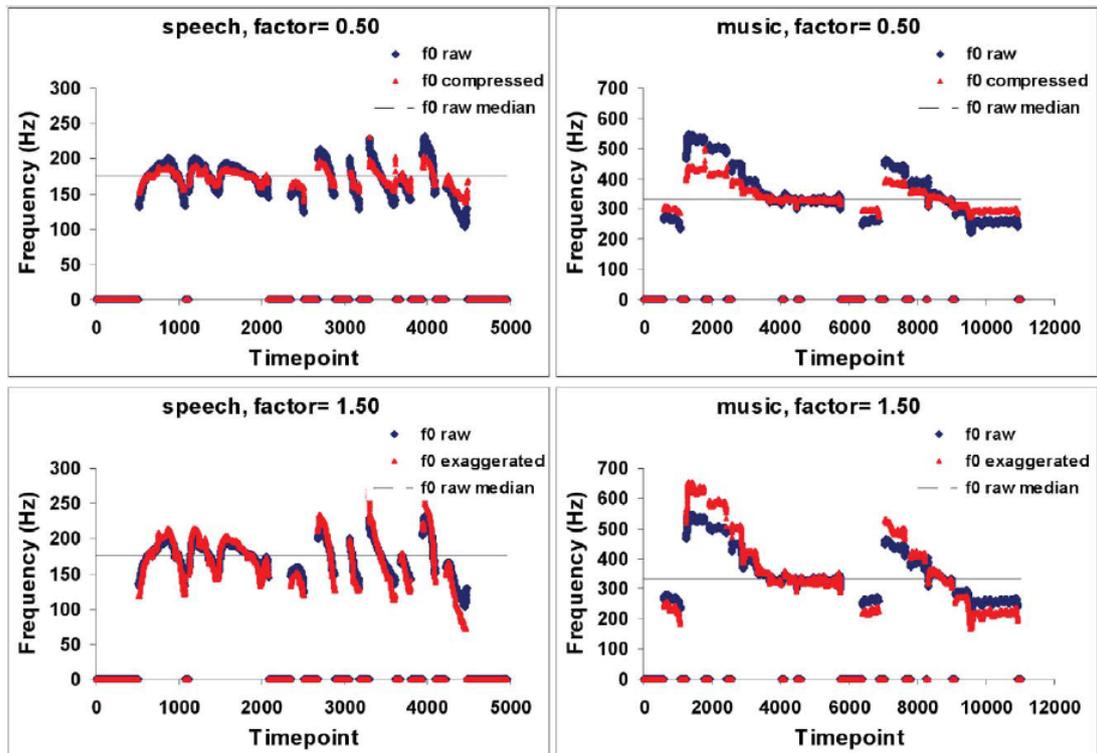


図 2.4: 話声と歌声の F0 軌跡：中心に対して 50 % 圧縮（上段），中心より 50 % 拡張（下段）（文献 [36] より引用）

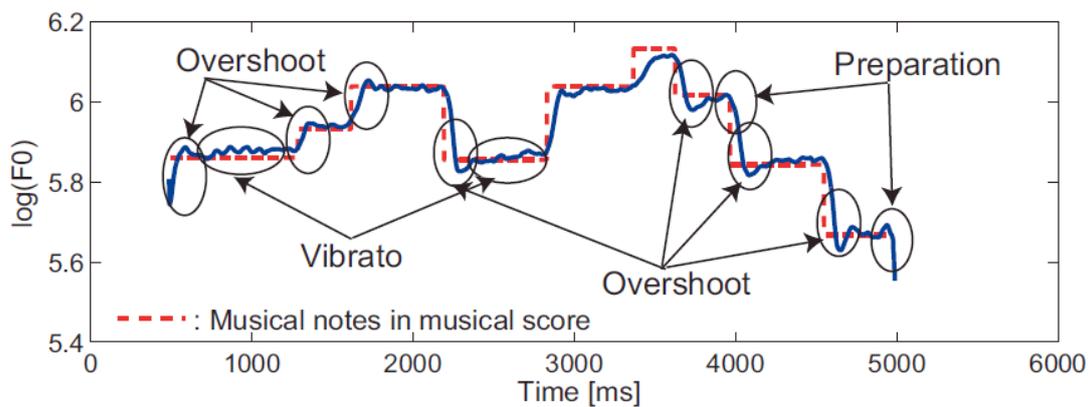


図 2.5: 歌声の F0 動的変動（ヴィブラート，オーバーシュート，プレパレーション）（文献 [43] より引用）

### 2.2.3 Singer's Formant

音響分析結果から，熟達した歌唱者の歌声は話声と異なるスペクトル形状を持つことが分かっている [31,44]．前述した大石らの検討において，朗読音声/歌声の自動識別時にスペクトル包絡時間変化の利用が有効であることが示されている [35]．この結果から，スペクトル形状が歌声を特徴づける性質であると言える．歌声に特有の 3 kHz 付近に観察されるスペクトルピークは，Singer's Formant と呼ばれる．Singer's Formant は，第 3～第 5 フォルマントの間隔が狭くなることで，この付近の周波数成分における音声の利得が増加することにより創出されると指摘されている [20]．齋藤らの検討から，この Singer's Formant が歌声の「響き」の印象と対応していることが明らかとなっている [39]．齋藤らによる，話声・歌声のスペクトル包絡の概形を図 2.6 に示す．Deutsch により，歌声とオペラ楽音のスペクトル包絡の関係が述べられている．ここでは，歌声独特のスペクトルピークが存在する部分において，楽音のみの包絡よりも歌声が加わった包絡の方が相対的に大きなエネルギーを持つことが示されている [45]．よって，Singer's Formant は楽音にマスキングされずに歌唱を知覚できるための重要な物理的性質であると言える．また，Singer's Formant は女性に比べて男性のほうが顕著であることが知られている [20]．

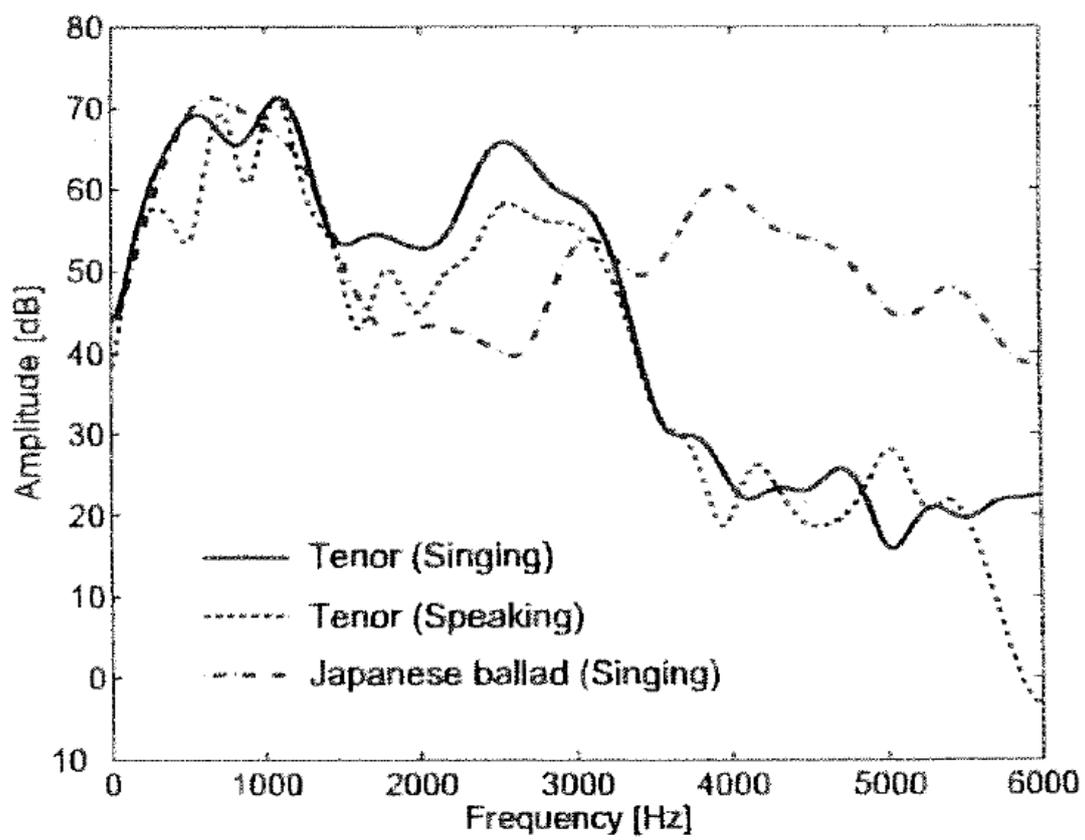


図 2.6: 話声・歌声のスペクトル包絡の概形 (文献 [39] より引用)

## 2.3 歌声の音響的特徴のまとめ

2.1 および 2.2 節で述べた知見を踏まえると、話声と歌声を対比させた際の音響的特徴の違いは、表 2.1 のようにまとめられる。これらの知見から、歌唱は音韻持続時間や発声強度の時間変化、F0 とその動的変動、Singer's Formant 等のスペクトル特徴という面で独特の音響的性質を有していることが分かった。特に、ヴィブラート等の F0 動的変動や Singer's Formant は「歌声らしさ」を特徴づける重要な音響的性質であると分かった。

表 2.1 から、話声と歌声の違いを表す特徴として、特に F0 に関連するものが多いことが分かる。F0 は音高（ピッチ）に対応する物理量であり、調子外れに対しても大きく関係する特徴であると考えられる。したがって、歌唱における聴覚フィードバックに着目するにあたり、音響的特徴として特に F0 に着目する必要があると考えられる。次章では、聴覚フィードバックに関連する現象および手法について述べるとともに、主に歌声の F0 が聴覚フィードバックとどのように関わっているのかについて、関連研究から得られた知見を述べる。

表 2.1: 話声と歌声の音響的特徴の違い

	話声	歌声
強さ	弱	強
持続時間	短	長
F0 (軌跡)	低	高
	連続的	離散的
		複数の声区
		独特の動的変動
スペクトル		歌声独特の スペクトルピーク (歌唱フォルマント)

## 第3章 聴覚フィードバックの関連研究

### 3.1 聴覚フィードバックに関連する現象と手法

ことばの鎖 [16] の考え方に基づくると、話者は相手の音声や外界の音を知覚し、発話のプランニングを行ない、自らの音声を生成する。この際、生成された音声は聴覚フィードバックとして話者自身に知覚され、音声生成系を制御するための情報として用いられる。話者内部における生成と知覚の関係を、図 3.1 に示す。音声生成と音声知覚の間の相互関係を深く調査することを目的として、これまで気導の聴覚フィードバックに着目した検討が行なわれている。本節では、聴覚フィードバックについての関連研究を調査し、聴覚フィードバックに関連する現象や手法についてこれまでに分かっている知見を述べる。

#### 3.1.1 ロンバード効果

Lombard により、発話中に騒音レベルが上昇して話者自身の音声の聴き取りにくくなる時、話者の音声レベルや  $F_0$ 、第一フォルマント周波数 ( $F_1$ ) などが上昇する現象が発見された [46]。この現象はロンバード効果 (Lombard Effect) と呼ばれる [47]。ロンバード効果は、周囲の騒音レベルの変化に応じて話者自身の音声を調整することで、他者とのコミュニケーションを円滑に保つようするための不随意的な現象であり、聴覚フィードバックの正帰還としての性質を示していると考えられている [18]。実際に、自動車雑音とバブル雑音を話者に呈示した際の、生成音声への影響が調べられている。この結果として、 $F_0$  の上昇と発話持続時間の増加が見られるだけでなく、呈示雑音が異なると音声のスペクトル包絡のピーク位置が異なることが示されている [48]。この結果は、ロンバード効果が単に聴覚フィードバックの知覚レベル変化に対する反応だけではなく、周囲の音環境に応じた多様な音声変化を伴うことを示している。また、ロンバード効果は、単に文章を読み上げるよりも、実際に他者とコミュニケーションをしている環境の方がその効果が大きいことも分かっている [49]。

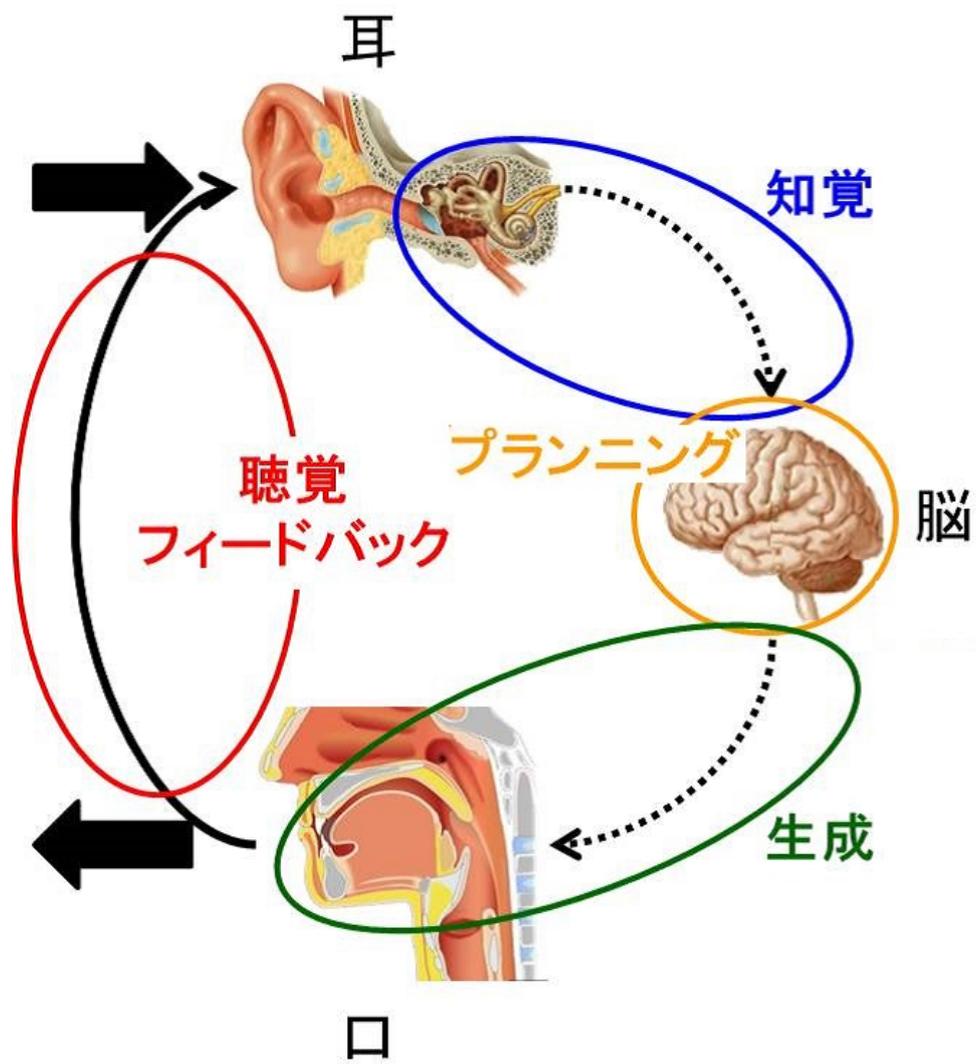


図 3.1: 話者内部における音声知覚と音声生成の関係

### 3.1.2 遅延聴覚フィードバック (DAF)

話者の音声知覚と音声生成との関係を示す実験的検討の1つに、話者が生成した音声を遅延させて、話者自身に知覚させる方法がある。この方法は遅延聴覚フィードバック (Delayed Auditory Feedback, DAF) と呼ばれる [47]。健常話者に対して、生成音声を 100–200 ms 程度遅延させて呈示すると、吃音に類似した発話スタイルや話速の減少が確認されることが知られている [50, 51]。この結果から、遅延により話者の音声生成-音声知覚間で時間的な整合性が破綻すると、正常な発話ができないことが分かる。すなわち、聴覚フィードバックにより生成-知覚がリアルタイムに作用し合っていると見える。話者が意図的に異なる話速 (普通/速い) で発話した際の DAF による影響を調査した Stuart らの検討でも、話者の話速に依らず 50–200 ms の遅延条件において部分的な音韻長の増加や繰り返しが多く観測され、話速も低下したという結果が示されている [52]。DAF 下で見られる発話への影響には性差があることや、長く訓練をすれば DAF 下においても流暢に発話ができるようになることも知られている [53]。

前述した知見とは反対に、吃音話者に対して 93 ms 程度の遅延で DAF を行なうと、話者の吃音症状が低減する傾向があることが報告されている [54]。この知見を応用して、実際に吃音話者に対するリハビリを目的として DAF の手法が用いられている [4]。

DAF の手法は比較的簡易な実験装置を用いて実現でき、これにより時間に関連する音声のパラメータに着目した知覚-生成間の相互関係について調査することができる。しかし、この手法は発話過程の破綻を伴うものであるため、得られた知見が知覚と生成の相互作用を理解する上で直ちに有用なものではないという可能性も指摘されている [18]。

### 3.1.3 変形聴覚フィードバック (TAF)

発話過程を破綻させず、生成と知覚のリアルタイムな関係性を検証する手法として、話者が生成した音声の音響パラメータをリアルタイムに微小変形した刺激を知覚させる方法がある。この方法は変形聴覚フィードバック (Transformed Auditory Feedback, TAF) と呼ばれる [55]。TAF の実験方法の概略図を図 3.2 に示す。

音響パラメータとして F0 に着目した Burnett らによる検討では、母音/a/を発声した話者の音声の F0 を 100 cent だけ上昇/下降させる摂動を与えたときの、生成音声への影響が調査された。その結果、変形とは逆方向に F0 が変化する補償反応が多く見られた一方で、一部の被験者では変形と同方向に F0 が変化する追従反応も見られた [56]。F0 の上昇刺激に対する補償反応の例を図 3.3 に示す。呈示音声の変形開始から生成音声に反応が出るまでの潜時 (遅れ時間) は、補償反応の際に平均 192 ms、追従反応の際に平均 327 ms であり、これらの間に有意差があることが明らかとなった。さらに、与える F0 摂動の大きさを 25–300 ms の間で変化させると、変形を大きくするほど補償反応を示す割合が減少し、追従反応を占める割合が増加する傾向が見られた。この結果から、小さく瞬時的な F0 変化に対する補償反応と、大きくゆったりとした F0 変化に対する追従反応の 2 種類が見られることが示唆された。Larson らにより、与える F0 摂動の変形速度を 100–10000 cent/s

の間で変化させた際の生成への影響が調べられ、摂動の変形速度が大きいほど、補償反応はゆっくりとした変化となる傾向が示された [57]。TAF による F0 変形刺激による生成音声への影響には性差や年齢効果が認められることも指摘されている [58,59]。

Larson らによる別の検討では、母音/u/を発声した話者の音声の F0・振幅に対してそれぞれ、もしくは同時に摂動を与えたときの、生成音声への影響が調査された。この結果、F0 の摂動に対しては F0 の補償反応が、振幅の摂動に対しては振幅の補償反応がそれぞれ独立に見られる傾向があり、各々の音響パラメータに対して独立に補償動作が働くことが示唆された [60]。

母音発声時に F0 摂動を与える TAF の下で fMRI を用いた脳活動計測を行なうと、右半球においては縁上回、前頭前野、前島、上側頭部および頭頂回溝に有意な活性が確認された一方、左半球においては運動前野のみに有意な活性が確認された [61]。特に、右前頭前皮質は音声の音高（ピッチ）識別や聴覚的作業記憶に関連して活動することが報告されており [62,63]、発声中に自らの音声からピッチを知覚・保持し、F0 変化に対応して生成系を制御することが脳科学の知見からも示唆されている。

連続音節刺激を用いた Donath らの検討において、100 cent 程度の F0 下降変形中に 157 ms 程度の遅れを伴って 40 cent 程度の F0 上昇が確認された [64]。また、F0 下降変形の終了後も 171 ms 程度変化が持続することが確認された。さらに、刺激として話声と持続母音の両方を用いた Chen らの検討によっても、話声刺激、持続母音刺激の両方において、50–200 cent の F0 摂動に対して補償反応が確認された [65]。これらの結果から、聴覚フィードバックによる音節内のリアルタイムな F0 のモニタリングだけではなく、言語発話中の長期的な F0 の補償動作の持続があることが示唆された。

音声のスペクトル変化として第一フォルマント（F1）の変化に着目した検討 [66] および第二フォルマントの変化に着目した検討 [67] も行なわれており、F0 摂動に対する補償反応と同様に、F1 や F2 への摂動に対しても補償反応が見られることが示されている。フォルマントへの摂動に対する補償動作は、音響分析結果からだけでなく、筋電信号の解析結果からも示されている [68]。

TAF の手法において注意すべき点として、話者が変形を受けない自身の骨導音声も知覚してしまうことがある。変形を受けない骨導音声をマスキングするため、刺激音声とともに、低域強調された白色雑音やピンク雑音が音圧レベル 50–90 dB 程度で呈示されている [55–57]。一方、前述したロンバード効果のように、雑音の呈示自体が話者の発話スタイルに影響してしまう可能性があることも指摘されており [56]、骨導音声の知覚や雑音呈示については議論の余地がある。

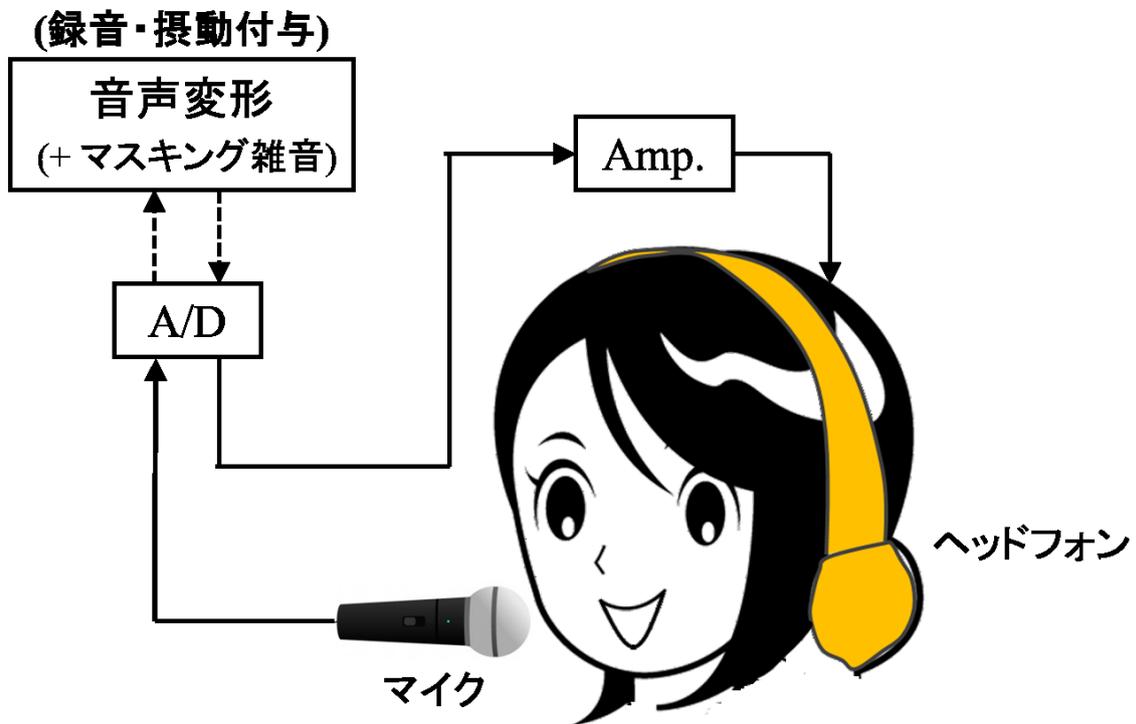


図 3.2: 変形聴覚フィードバック (TAF) の実験方法の概略

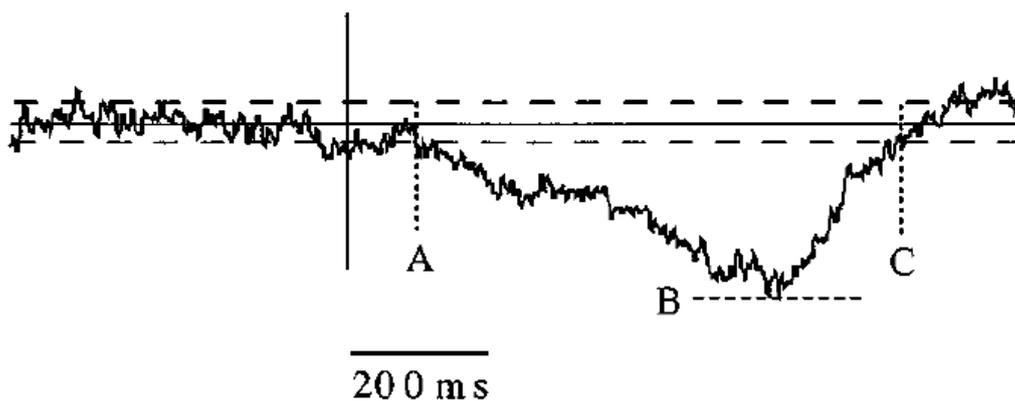


図 3.3: TAF による F0 の上昇刺激に対する補償反応の例 : (A) 反応の始点, (B) 反応のピーク, (C) 反応の終点 (文献 [56] より引用)

## 3.2 歌声を対象とした関連研究

前節での知見から，聴覚フィードバックには発話中に自らの音声の音響パラメータをリアルタイムに補正する働きがあることが分かった．したがって，歌唱時の聴覚フィードバックにも同様の働きがあると予想できる．しかし，歌声は第2章で述べた歌唱独特の音響的特徴を有しており，前節の知見だけでは歌唱時の聴覚フィードバックについて深く議論することはできない．そこで本節では，歌唱を対象とした聴覚フィードバックに関する関連研究を調査する．

歌声を対象とした聴覚フィードバックの関連研究では，主に次の二つのアプローチにより，歌声のF0がどのように変化するかについて検討されている．

- 聴覚フィードバックの変形
- 話者内部の手がかりへの着目（聴覚フィードバックの遮断）

ここでは，それぞれのアプローチによる関連研究について述べ，得られた知見と問題点を整理する．

### 3.2.1 聴覚フィードバックの変形

歌唱に関連した発声課題として，ターゲット信号にピッチを合わせて追唱させるものがあり，ピッチマッチング (pitch-matching) [69, 70] と呼ばれる．ピッチマッチングの精度は音楽的訓練により向上することが示唆されており [69]，ピッチマッチング時にはターゲット信号として人工合成音を用いるよりも実際に録音された歌唱音声を用いた方が精度が高いという結果を示した知見 [70] もある．

歌唱時の聴覚フィードバックに関する研究のアプローチの一つとして，歌声に微小変形を施して気導音声として呈示し，ピッチマッチングの精度を調べる方法が行なわれている．歌声の音程に食い違いが生じた際のF0の修正に着目した Keough & Jones の検討では，歌唱者の音声を予め録音してターゲット信号とし，被験者にはターゲット信号と等しいピッチを保って歌唱させる課題が行なわれた．その際，ターゲット信号を上または下に 2 cent (2 半音) 変形させる摂動試行を入れ，歌唱熟達者・非熟達者の歌声のF0の推移を調べた．その結果，両者ともに，摂動に対して補償しながら歌唱を行なった．歌唱熟達者は非熟達者よりも補償反応時のF0中央値とF0平均との差が小さく，歌唱熟達者はより精密にF0の食い違いを補償していることが明らかとなった [71]．Keough & Jones の検討で得られた，歌唱者のF0中央値とF0平均の結果を図3.4に示す．この知見は，聴覚フィードバックに基づくF0補正能力が存在し，またそれが歌唱経験により強化されることを示している．その他，歌唱者の歌声に高域強調やF0の変形を施したときには，全体的に歌声のF0が下降する傾向があることが報告されている [72]．なお，これらの変形により歌声のF0が下降する反応が生じるメカニズムについては，詳細には述べられていない．

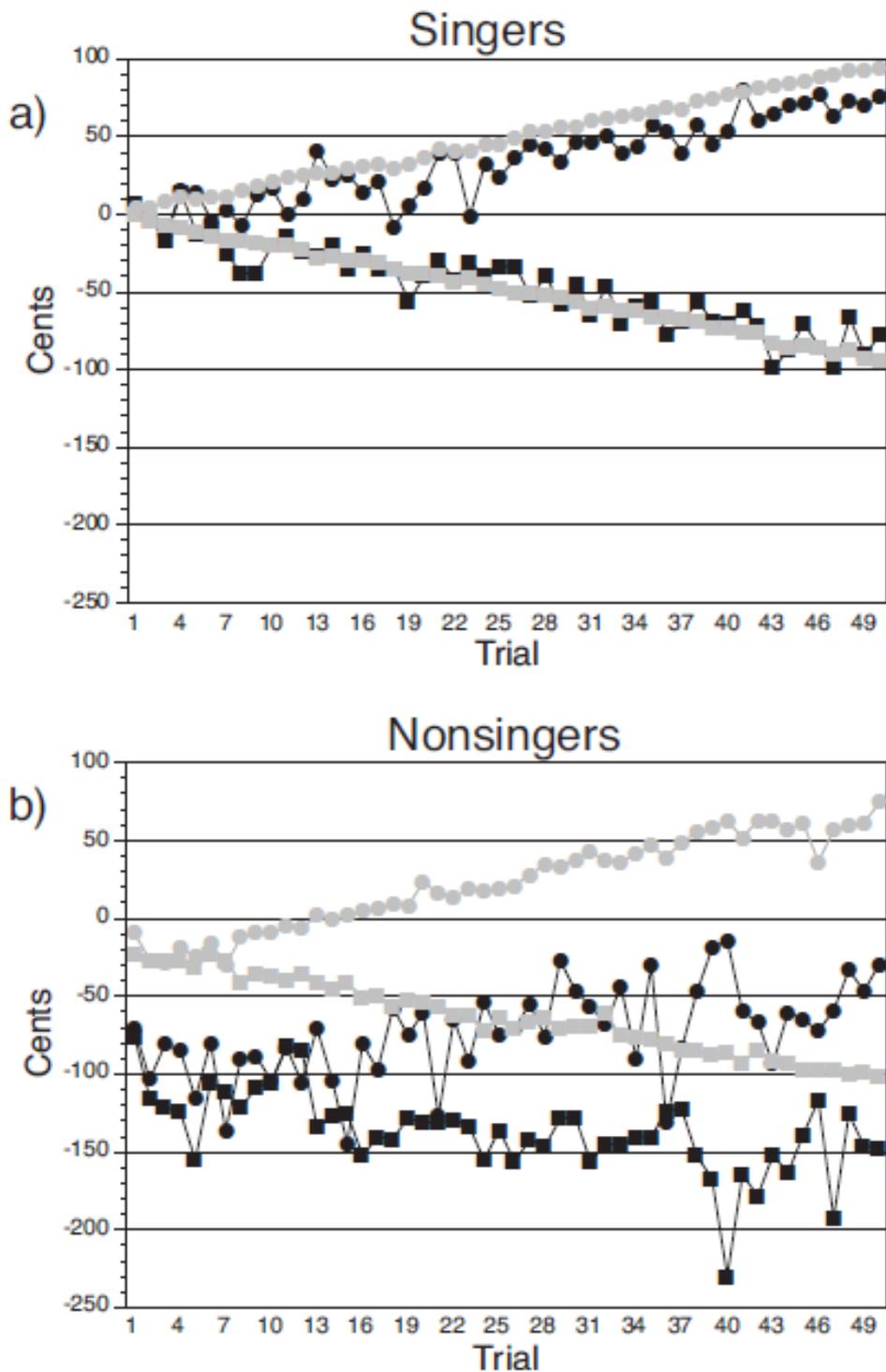


図 3.4: F0 上昇/下降ターゲット信号への追唱課題に対する歌声の F0 推移 : F0 の試行内平均 (灰色) , F0 の試行内中央値 (黒色) (文献 [71] より引用)

歌声のヴィブラートと聴覚フィードバックの関係を明らかにするため、DAFの下で歌唱を行なうことで知覚-生成の間でヴィブラートの時間的整合を破綻させて生成音声の影響を調べる検討もある。この検討では、100–500 msの遅延を伴うDAF下にて、遅延時間が増加するほどヴィブラートの速さが減少するという結果や、300 msの遅延を伴うDAF下にて、通常のヴィブラートに加えゆっくりとしたF0の周期的変動が見られるという結果が報告されている [20]。しかし、これらの結果からは、歌声のヴィブラートと聴覚フィードバックの間の詳細な関係を議論するまでには至っていない。

その他、歌唱者の音声に周波数変調を施して歌唱者自身に知覚させる TAF の試みが行なわれている。Leydon らは、歌唱者の音声の F0 に対して  $\pm 25$  cent の周期的変動を付与することでヴィブラートを模擬し、変調周波数 1–10 Hz の条件で歌唱者自身に知覚させて生成音声への影響を調べた。この結果、生成音声の F0 にも平均 5 Hz の変調成分が確認された [73]。利得・位相特性の分析から、フィードバック音声と生成音声の間の F0 変調成分が同相の際にこの利得がピークとなる傾向が示された。この結果は、歌唱時のヴィブラートの創出に聴覚フィードバックが大きく関係していることを示す。しかし、この検討における生成音声の反応が、3.1 節内で述べた聴覚フィードバックによる補償動作に基づく結果であるのか、もしくは追従動作に基づく結果であるのかについては詳しく調査されていない。Leydon らの検討において予想された、目標とする F0 軌跡、呈示される TAF の音声の F0 軌跡、生成される歌声の F0 軌跡との関係を表した概略図は、図 3.5 のようになる。

前項における TAF の手法と同様に、歌唱における聴覚フィードバックの変形に関する検討でも、音圧レベル 80 dB 程度のバブル雑音を呈示したり [71]、刺激音声を発声レベルに対して 10 dB ほど増幅させて呈示したり [73] することで骨導音声のマスキングを行なっている。

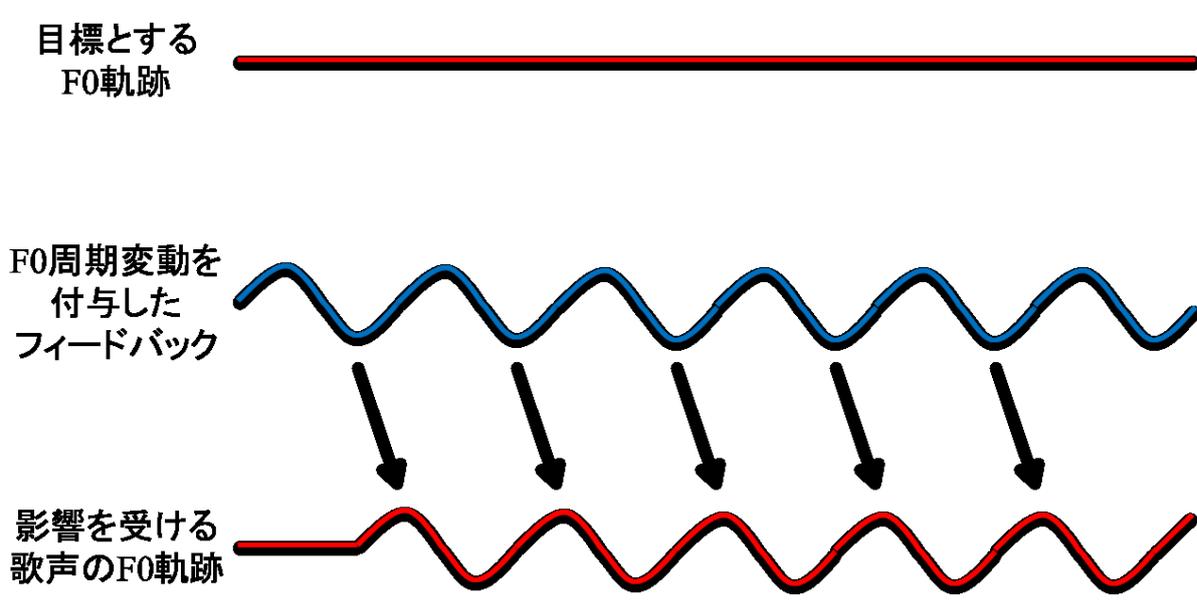


図 3.5: F0 の変調刺激と生成音声の F0 軌跡との関係の概略 (文献 [73] をもとに作成)

### 3.2.2 話者内部の手がかりへの着目

歌唱時の聴覚フィードバックに関する研究のもう一つのアプローチとして、気導音声とは異なる話者内部の手がかり着目するため雑音を気導呈示して歌唱者自身の音声をマスクした下で、歌声の F0 の変化を調べる方法が行なわれている。

歌唱熟達者を対象とした Mürbe らによる検討では、旋律の音程に合わせた母音/a/によるアカペラ歌唱時に、50–2000 Hz に帯域制限された白色雑音を音圧レベル 105 dB で呈示した際の生成音声の F0 の変化が調査された。このとき、発声の持続時間の条件として legato (切らない)・staccato (切る)、速さの条件として slow (ゆっくり)・fast (速い) が設定された。この結果、持続時間と速さに対するすべての条件において、雑音呈示時の方が非呈示時に比べ旋律の音程からの F0 逸脱が大きくなった [74]。この結果は、意図したピッチに合わせて歌声の F0 を制御する際に、聴覚フィードバックを用いた歌声のモニタリングが行なわれることを示唆している。歌唱熟達者・非熟達者の両方を対象とした同様の実験も行なわれており、雑音非呈示時には両者の F0 逸脱にほとんど差がない一方、雑音呈示時には熟達者に比べ非熟達者の方が F0 逸脱が大きい結果が報告されている [20]。この結果をもとに、F0 制御時には歌唱熟達者よりも非熟達者のほうが聴覚フィードバックへ依存する傾向があることが指摘されている。

Watts らは、歌唱訓練者 (TS)・訓練なし熟達者 (UTS)・訓練なし非熟達者 (NTS) の 3 種類の被験者群を対象に、ターゲット信号へのピッチマッチング課題を行なった。この際、スピーカーからターゲット信号を呈示する条件 (EF 条件) と、ヘッドフォンからターゲット信号とともに音圧レベル 85 dB の白色雑音を呈示する条件 (IF 条件) が設定され、ターゲット信号の F0 に対する歌声の F0 逸脱が調べられた。その結果、被験者群の間で F0 逸脱の大きさを比較すると、歌唱に熟達しない NTS 群が最も大きかった [75]。しかし、被験者群の中で雑音呈示の有無 (EF/IF 条件の別) による結果の違いを比較すると、歌唱に熟達する TS・UTS 群では EF/IF 条件の違いによる F0 逸脱の差はあまり見られなかった一方で、NTS 群では雑音を呈示する IF 条件時のほうが、雑音を呈示しない EF 条件時に比べ F0 逸脱が小さかった。気導からの雑音呈示により歌唱者の聴覚フィードバックがマスクされているとするならば、Watts らによる結果からは歌声のピッチマッチングに聴覚フィードバックが必ずしも重要であるとは言えない。また、Watts らによって示されたピッチマッチングの精度の傾向は、TS 群では IF 条件に比べ EF 条件でわずかに高く、UTS 群ではその逆となっていた。この傾向をもとに、Watts らは、ピッチマッチング能力は生得性を有する一方、歌唱訓練は聴覚フィードバックに依存したピッチマッチング能力をトレーニングさせているのではないかと推測している。歌唱者のピッチマッチングの精度に関連する知見として、絶対音感がある。絶対音感は一般に、ピッチ情報を音楽的な音名と対応づけてカテゴリ化されることを指す [8]。絶対音感保有者では予めピッチ情報が長期的な記憶として保持されていると考えられ [76, 77]、非保有者とは異なりピッチの判断が高速に自動化されると解釈されている [78]。これらの知見を踏まえると、先の Watts ら [75] の結果は絶対音感等、話者内部に存在する手がかりをもとにした発話スタイルを表している可能性が示唆される。

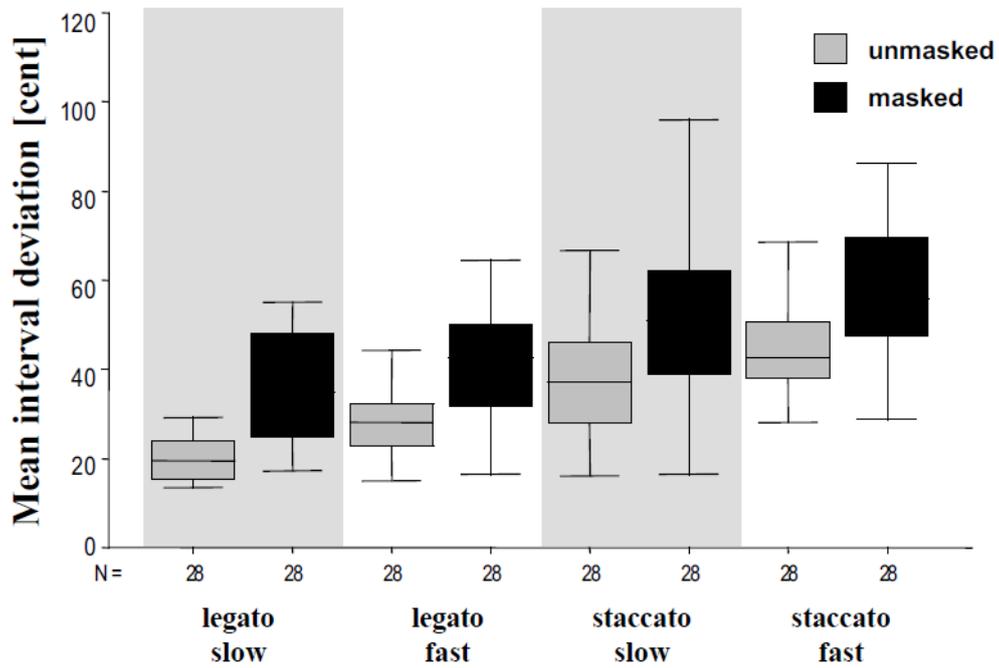


図 3.6: 雑音呈示/非呈示時における F0 逸脱の比較 (文献 [74] より引用)

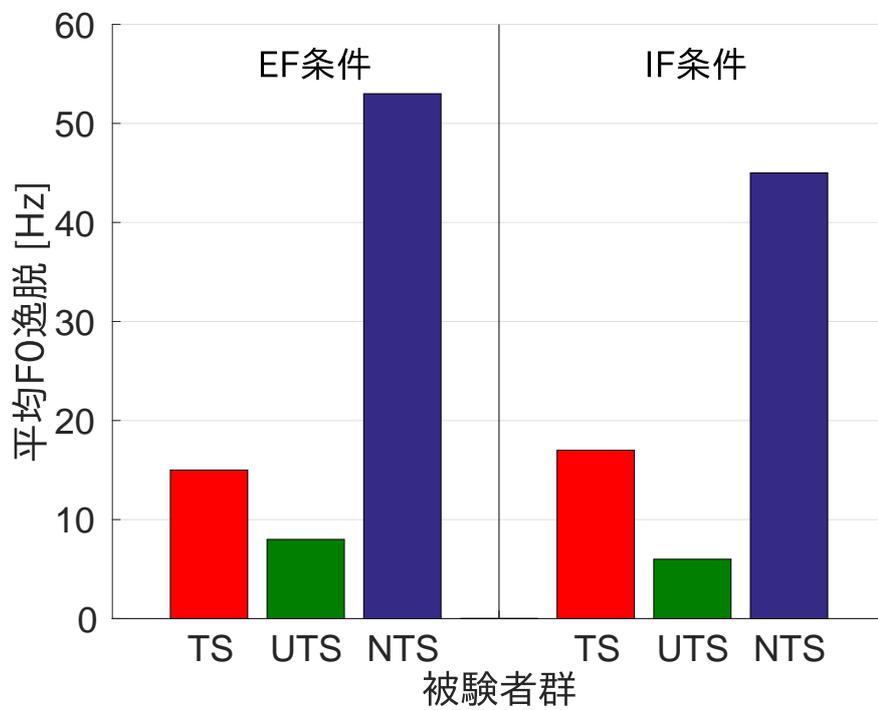


図 3.7: EF/IF 条件における平均 F0 逸脱の比較 (文献 [75] のデータをもとに作成)

### 3.3 これまでの知見のまとめ

#### 明らかになったこと

まず、聴覚フィードバックに関連する発話中の現象としてロンバード効果があり、周囲の環境に即した音声のモニタリングが行なわれることが分かった。また、知覚-生成間の相互作用に着目するための手法として、音声の時間情報に着目した遅延聴覚フィードバック (DAF) と、F0 やスペクトル情報などのパラメータに着目した変形聴覚フィードバック (TAF) があることが分かった。

次に、歌声を対象とした聴覚フィードバックの関連研究のアプローチの一つ目として、聴覚フィードバックの変形が歌声の F0 に与える影響が調査されていることが分かった。これらの検討から、次のことが明らかとなっていることが分かった。

- 聴覚フィードバックにより、ターゲット信号の F0 と歌声の F0 の間の食い違いを補償する。
- F0 の周期的変動を付与した TAF の下では、生成音声の F0 にも変動成分が見られる。

歌声を対象とした聴覚フィードバックの関連研究のもう一つのアプローチとして、雑音呈示により気導の聴覚フィードバックをマスキングした際の歌声への影響も調査されていることが分かった。これらの検討から、次のことが明らかとなっていることが分かった。

- アカペラ歌唱では、雑音呈示時に非呈示時よりも目標からの F0 逸脱が大きくなる。
- ターゲット信号に音程を合わせる課題では、雑音呈示時に非呈示時よりもピッチマッチングの精度が向上する傾向がある。

上記の知見から、聴覚フィードバックが歌声の F0 変化に対して重要な役割を果たしていることが予想される。一方、歌唱時には聴覚フィードバックのみに依存してはならず、話者内部に存在する手がかりにも依存していることが予想される。

#### 課題点

これまでは外部からの音声呈示によって、変形したフィードバック音声を気導音声として知覚させる、もしくは気導の聴覚フィードバックを遮断する方法が行なわれている。これらの方法は、歌唱における気導の聴覚フィードバックの性質や役割について調査することを前提としている。一方で、歌唱者自身が生成した歌声を自然に知覚するプロセスにおいて、気導音声と骨導音声の両方が存在する。骨導による聴覚フィードバックが、歌声特有の音響的特徴を伴う歌声生成にどのように寄与しているのかは、依然として未解明である。知覚・生成間の相互作用に対して、気導・骨導の聴覚フィードバックがどのように関係しているのかについて、検討する必要があると考えられる。さらに、自らの音声をマス

キングするための雑音の存在が生成音声にどのように影響しているのかについては、注意して検討すべき項目であると考えられる。

次章では、骨導の聴覚フィードバックの経路とその特徴について述べ、既存知見からの考察を図る。

## 第4章 気導・骨導の聴覚フィードバック

本章では、骨導の聴覚フィードバックの経路と音響的特徴、知覚的な性質について述べる。また、これらの知見をもとに、歌唱における聴覚フィードバックの性質について、気導・骨導という着眼点により整理する。

### 4.1 骨導音声の音響的特徴と知覚特性

ヒトが聴覚フィードバックによって自らの生成音声を知覚するときには、口外に放射された音声は空気を伝導して聴覚系に到達するもの（気導音声）だけでなく、音声は口外に出ずに体内で骨などを伝導して聴覚系に到達するもの（骨導音声）も存在する。Stenfeltによる生理学的な調査によって、骨導による音声の知覚経路は(1)外耳道内放射、(2)中耳耳小骨の慣性振動（慣性骨導）、(3)内耳リンパ液の慣性振動、(4)内耳リンパ液の圧縮（圧縮骨導）、(5)脳脊髄液の圧力伝達に分類されることが分かっている [79]。外耳道内放射では 2.7 kHz 付近がピークとなり、高域通過フィルタの特性がある一方、中耳耳小骨の慣性振動では、耳小骨の共振周波数である 1 ~ 3 kHz 付近の成分が強調されることが示されている [80]。また、内耳リンパ液の慣性振動では 1 kHz 以下の低域成分が [81]、圧縮骨導では 4 kHz 以上の高域成分が伝達されること [82] が、それぞれ指摘されている。さらに、脳脊髄液の圧力伝達は 20 Hz 以下の低域のみに寄与することが指摘されている [81]。なお、聴覚としての知覚ではないが、骨振動が皮膚触覚を介して 500 Hz 以下の振動知覚に寄与していることも指摘されている [81]。骨導音声の知覚経路と、伝達される周波数成分の関係を表 4.1 に示す。また、それぞれの伝達経路の概略図を図 4.1 に示す。

表 4.1: 骨導音声の知覚経路と寄与する周波数成分

到達部分	経路	周波数成分
外耳	(1) 外耳道内放射	2 ~ 3 kHz
中耳	(2) 耳小骨慣性振動 (慣性骨導)	1 ~ 3 kHz
内耳	(3) リンパ液慣性振動	~ 1 kHz
	(4) リンパ液圧縮 (圧縮骨導)	4 kHz ~
	(5) 脳脊髄液の圧力伝達	~ 20 Hz
(聴覚以外)	皮膚感覚	~ 500 Hz

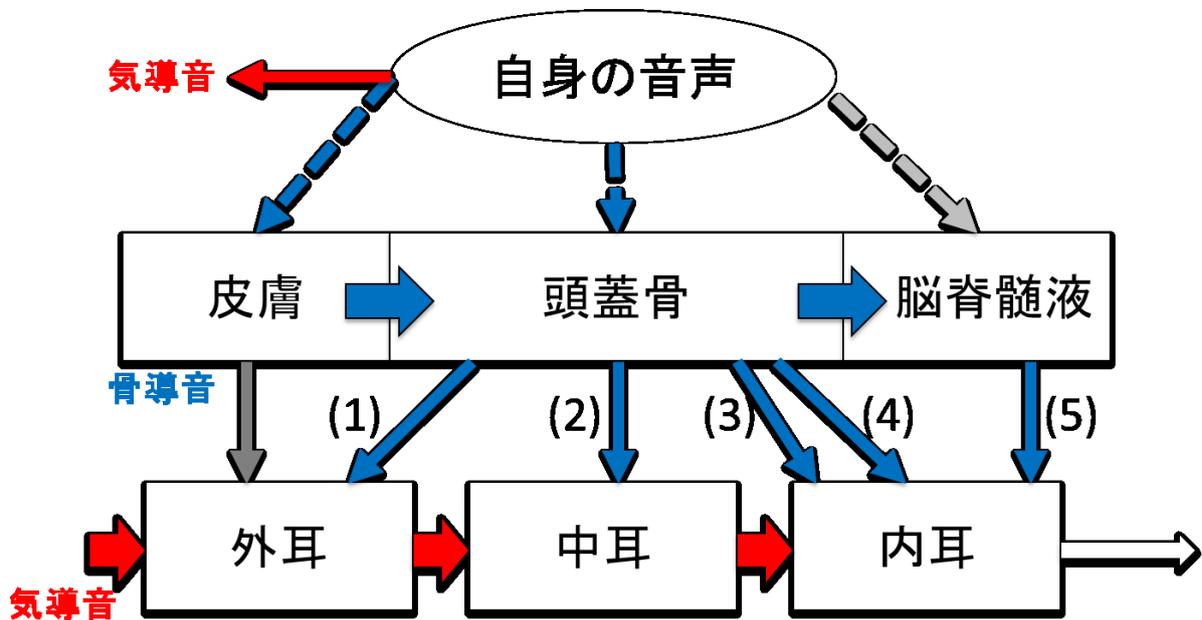


図 4.1: 骨導音声の伝達経路の概略 : (1) 外耳道内放射 , (2) 中耳耳小骨の慣性振動 , (3) 内耳リンパ液の慣性振動 , (4) 圧縮骨導 , (5) 脳脊髄液の振動 (文献 [79] をもとに作成)

マイクロホンから収録された気導音声と、骨導マイクから収録された骨導音声の音響分析されている。Won & Berger は、気導音声から骨導音声への伝達関数を求めることを目的として、歌声に対する気導・骨導音声の周波数特性を分析した。図 4.2 は、この結果を示したものである。この結果より、骨導の歌声は気導のものに比べて 4 kHz 以上の高域成分が低下する一方で、F0 付近の成分は尖鋭となることが確認できる [83]。話声を用いた Tamiya & Shimamura による分析結果からも、骨導音声は 1 kHz 以下の周波数成分を多く含む一方で、気導音声に比べて高域成分が低下していることが確認できる [84]。骨導音声の音響的特徴として高域成分が低下するという傾向は複数の検討から示されているが、Kondo らによりこの高域低下の程度には収録機器依存、および話者依存があることが指摘されている [85]。この検討では、低下される高域成分を回復させることにより、骨導音声の音質改善に一定の効果が得られることを示唆している。さらには、骨導音声収録機器の装着位置が、音声了解性と音質に影響を及ぼすことも指摘されており、了解性・音質のどちらも前額部へのマイク装着時の骨導音声において最も高いことが分かっている [86]。

自らの音声を知覚する際の気導・骨導の優位性について、知覚的側面に基づいて調査されている。Reinfeldt らは、10 個の母音発声においてイヤーマフ装着/非装着時の外耳道音圧レベルを測定することにより気導音声に対する骨導音声の音圧レベル比を測定した。さらに、スピーカーおよび骨導トランスデューサからの呈示音について、気導音声に対する骨導音声の最小可聴値の比を測定した。彼らは、これら音圧レベル比と最小可聴値の比の大小関係に基づき、気導音声と骨導音声のどちらの知覚が支配的であるかを周波数の関数として分析した。その結果、発声された母音の間で結果に差はあったが、あらゆる母音発声で 1~2 kHz における骨導音声の知覚が支配的であったことを報告している [87]。別な検討ではマスキングの手法が用いられ、骨導音声は 0.7~1.2 kHz における成分の知覚に、気導音声はそれ以外での成分の知覚に対して支配的であるということが報告されている [81]。

気導音声は話者の外部から聴覚系に到達するのに対し、骨導音声は話者内部を伝わる感覚的な手がかりであると言える。話者内部を伝わる感覚情報に関する関連研究として、喉頭振動の触覚フィードバックに着目した検討がある。Wang らは、母音/a/の発声時に、聴覚刺激と喉頭振動刺激に対して F0 摂動を与えた際の生成音声の F0 への影響を調査した。このとき、聴覚刺激はヘッドフォンから、喉頭振動刺激は喉頭部に固定された骨導トランスデューサから呈示された。結果として、聴覚刺激・振動刺激時ともに、摂動とは反対方向への補償反応が多く見られ、喉頭振動と聴覚の両方での刺激時に最も反応が大きかった。さらに、振動刺激は聴覚刺激に比べて低い潜時での補償反応を誘発した [88]。この結果から、発話における話者内部の感覚情報として、話者自身の喉頭振動の知覚が大きな役割を果たしていることが示唆される。

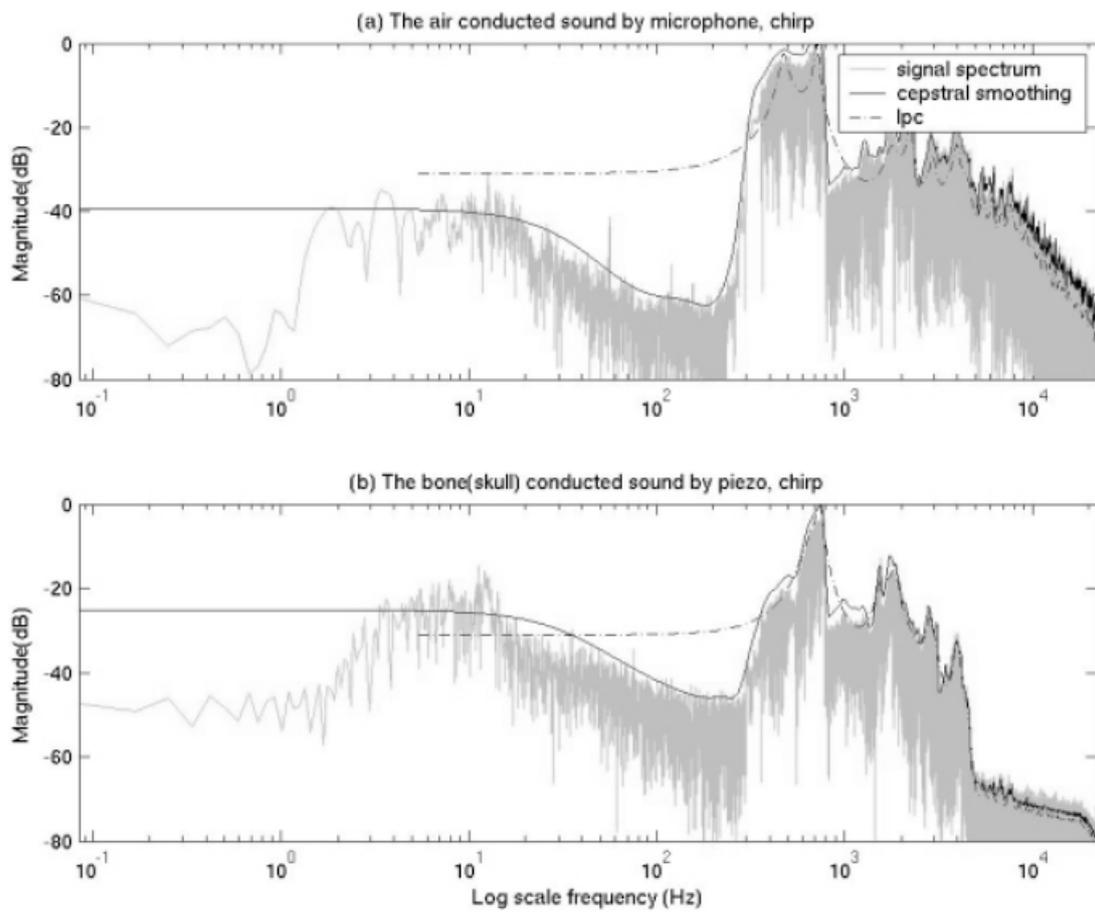


図 4.2: 気導・骨導音声の周波数特性の比較 (文献 [83] より引用)

## 4.2 気導・骨導による歌声の特徴伝達に関する考察

最初に、歌唱における気導・骨導および喉頭振動のフィードバックについて考察する。F0 付近の成分を含む気導・骨導音声の周波数特性（図 4.2）が、歌唱者が実際に知覚する気導・骨導音声の性質を反映したものであるならば、気導・骨導の聴覚フィードバックの両者が歌声の F0 に関係していると考えられる。気導音声・骨導音声・喉頭振動を含めたフィードバックによる F0 制御の概略図として、図 4.3 を考える。図 4.3 のように、歌唱時の F0 の制御に対して、気導・骨導・喉頭振動のそれぞれがフィードバックとして機能していると仮定すると、このうちいずれかのフィードバックに不具合が生じたとしても、残りのフィードバックの働きにより F0 の制御は継続されると考えられる。したがって、前章における Watts ら [75] の検討では、歌唱時に気導の聴覚フィードバックを用いることができないとしても、話者内部を伝わるフィードバックを用いてより高精度にピッチマッチングができたと推測される。

次に、それぞれのフィードバックの特性と優位性について考察する。先の Wang らの知見 [88] を骨導の聴覚フィードバックにも拡張すると、気導・骨導の聴覚フィードバックは互いに異なる特性（利得、潜時など）を持つ可能性があると考えられる。特に、喉頭振動刺激の摂動に対する短い潜時での補償反応から、骨導の聴覚フィードバックは、F0 情報の伝達が気導によるものに比べて速いと推測できる。このことを仮定した場合、周期の短いヴィブラートや短時間で変動するオーバーシュート等の動的変動には、より伝達の速い骨導の聴覚フィードバックや、喉頭振動のフィードバックが寄与しているのではないかと推測することができる。また、骨導音声の音響的特徴が話者により異なること [85] を踏まえると、骨導の聴覚フィードバックの性質は話者によって異なっている可能性も考えられる。したがって、歌唱におけるピッチマッチングの精度には、それぞれのフィードバックの優位性も大きく関係しているかもしれない。

以上より、これまでに未考慮であった骨導の聴覚フィードバックには、歌声の F0 やその動的変動に対して影響があるものと考えられる。Sundberg は、気導の聴覚に対して補完的に働く他のフィードバック信号を見つけることが有効であることを述べており [20]、また先の Wang らの検討結果からは喉頭振動の触覚知覚がその他のフィードバックの 1 つとして働くことが示唆される [88]。本稿の調査結果から、骨導の聴覚フィードバックもまた、歌唱における重要なフィードバックの 1 つであると考えられる。

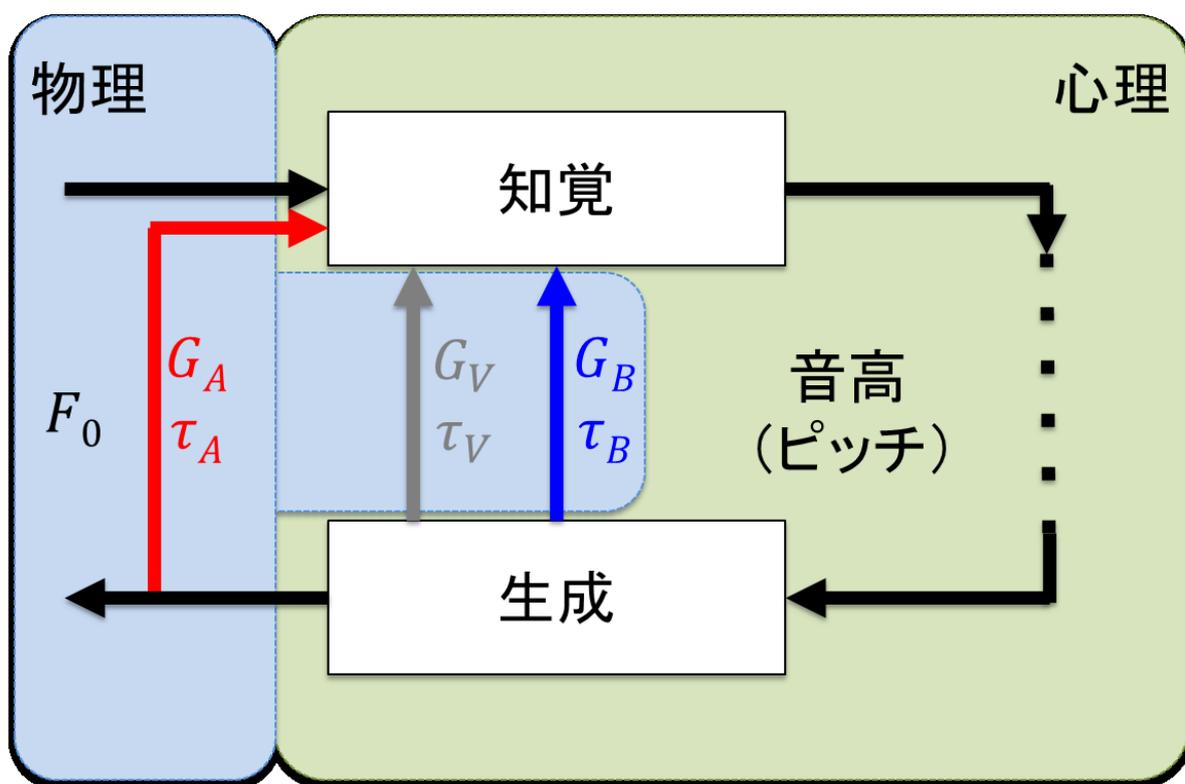


図 4.3: 気導・骨導・喉頭振動のフィードバックによる F0 制御の概略: 気導の聴覚フィードバックの利得および潜時 ( $G_A, \tau_a$ ), 骨導による聴覚フィードバックの利得および潜時 ( $G_B, \tau_B$ ), 喉頭振動によるフィードバックの利得および潜時 ( $G_v, \tau_v$ )

### 4.3 気導・骨導の聴覚フィードバックの性質を明らかにするための指針

前節における考察から，歌唱における気導・骨導それぞれの聴覚フィードバックの性質について下記のような予測を立てた．

- 歌声の F0 に対して，気導だけでなく骨導の聴覚フィードバックが大きく寄与している．
- 気導や骨導の聴覚フィードバックの優位性は，歌唱者により異なる．

これらのことを実証するためには，気導音声に着目された聴覚フィードバックに関する知見を骨導音声に対しても拡張して，気導・骨導音声は歌声に与える影響の違いを調べる必要がある．実証に向けた予備検討として，気導・骨導音声を用いて DAF を行ない，話者の発話スタイルにどのような変化が見られるかを調査した．予備検討の概要・実験方法と結果を，付録に示す．予備検討の結果，話者の発話スタイルの変化は，骨導音声による遅延音声の呈示において顕著に見られる傾向があった．したがって，歌声に F0 に対しても，骨導の聴覚フィードバックが歌唱に大きく寄与しているのではないかと考えられる．上記の実証のため，下記に着目した検討が必要となると考えられる．

#### 気導音声・骨導音声の音響的特徴の違いの精査

気導・骨導それぞれの聴覚フィードバックに着目するためには，気導音声と骨導音声の両方としての収録・呈示が不可欠となる．しかし，4.1 節において述べた通り，骨導音声の音響分析結果は，収録・呈示機器に依存する特性が含まれている可能性があり，音声自体の特性とは大きく異なると考えられる．この問題を解消するためには，呈示・収録に用いられる機器の特性を計測した上で，その特性を排除するためのフィルタ構築などの工夫が必要である．この検討により，図 4.3 の青色の部分に対応する物理的特徴を明らかにすることができると思われる．

#### 気導音声・骨導音声の知覚的な違いに関する検討

F0 情報が等しくてもスペクトル包絡情報が異なる 2 つの音では，知覚されるピッチの違いが生じることが示されている [89]．等しい F0 を持つ気導・骨導音声から知覚されるピッチに違いがあるとすれば，気導・骨導の聴覚フィードバックの働きが異なるという可能性もある．このことを調べるためには，気導・骨導音声のピッチ知覚について検討する必要がある．また，気導・骨導音声から，Singer's Formant に起因する知覚的作用があるのかどうかについても，検討すべき項目であると考えられる．さらに，骨導音声知覚に関する一般的な性質を明らかにするため，山本 [90] の指摘を参考に，骨導音声の位相

特性に着目して骨導音声の位相変化に対する知覚変化を調べるのが有効である可能性がある。この検討により、図 4.3 の緑色の部分に対応する心理的性質を明らかにすることができると思われる。

## 気導・骨導音声の振動呈示が歌声に与える影響の検討

気導・骨導音声のフィードバックとしての働きを知るためには、気導音声を用いた研究での手法と同様にして、TAF による振動刺激の呈示が歌声に与える影響を、気導・骨導音声の両方を用いて調べる必要がある。この際、歌唱者の自然な音声を、気導音声・骨導音声で独立して操作することが困難であるという問題がある。気導音声を用いた研究の大多数では、40～70 dB 程度の白色雑音やピンク雑音を呈示し、話者の自然な音声のマスキングを試みていた。しかしこの手法では、話者の気導音声もしくは骨導音声の一方を選択的にマスキングすることは困難であると思われる。

一つの方法として、von Békésy による気導音-骨導音相殺実験の拡張がある、この検討では 400 Hz の正弦波を用い、骨導音に対して位相を反転させ気導音として呈示することで互いに相殺することに成功した [91]。本研究では音声のリアルタイムな信号処理を必要とするため実装は困難である可能性があるが、歌唱者の骨導音声に対して気導から音声を逆位相呈示することができれば、骨導音声の知覚を外部から操作することが可能となると考えられる。

## 第5章 結論

本論文では、歌声に見られる音響的特徴の調査、発話や歌唱における気導の聴覚フィードバックに関する調査と、骨導音声の知覚や音響的性質に関する調査を行なった。その調査結果を下記に要約する。

### 歌声の音響的特徴

歌声は話声よりも長い音韻持続時間、強い音声パワーが見られることが分かった。また、歌声特有の音響的特徴として F0 軌跡と F0 動的変動（ヴィブラート等）、Singer's Formant があることが分かった。調子外れに関連する特徴として、F0 軌跡と F0 動的変動が重要であることが分かった。

### 聴覚フィードバックの関連研究

歌唱における F0 の追唱と F0 動的変動の創出に対して気導の聴覚フィードバックが影響していることが分かった。一方で、上記の特徴の創出に対して骨導の聴覚フィードバックも関係しているのかどうかは依然として未解明であることが分かった。

### 骨導音声の音響的特徴と知覚特性

骨導音声は主に 5 つの知覚経路を持つことが分かった。音響分析から、骨導音声には F0 付近の成分は十分に含まれる一方、4 kHz 以上の高域成分は低下している可能性があることが分かり、また、分析結果には個人差が大きいことも分かった。生理的・心理物理的検討から、骨導音声は 0.7~2 kHz 付近の周波数域の知覚に主に寄与していることが分かった。また、喉頭振動・聴覚の両方によるフィードバックは聴覚フィードバックのみの場合よりも強く、話者の F0 制御に影響していることが分かった。また、喉頭振動刺激による F0 の補償反応は聴覚刺激によるものよりも低い潜時を持つことも分かった。

結論として、上記の調査結果から、気導・骨導の聴覚フィードバックが歌声に与える影響を考察し、その予想を下記のように立てた。

- 歌声の  $F_0$  に対して，気導だけでなく骨導の聴覚フィードバックが大きく寄与している．
- 気導や骨導の聴覚フィードバックの優位性は，歌唱者により異なる．

今後はこれらの見通しを実証するため，骨導音声に関するデバイスの特性調査と，気導・骨導音声の知覚の検討，気導・骨導音声の振動呈示が歌声に与える影響について検討する予定である．

# 付録

## 気導・骨導音声を用いた遅延聴覚フィードバック

気導音声を用いた遅延聴覚フィードバック (DAF) の実験 [50] [52] により, 話者の音声を 100–200 ms 程度遅延させて話者自身に知覚させると発話時間の増加や吃りが見られることが分かっている. 一方で, 骨導からの遅延音声の話者の発話スタイルにどのような影響を与えるのかについては分かっていない. 本研究において気導・骨導の両方の聴覚フィードバックに着目するにあたり, 骨導音声を用いた DAF の実験を気導音声の手法と同様に行なうことで有益な知見が得られると考えた. そこで, 本研究の予備検討として, 気導・骨導音声のそれぞれから遅延音声を呈示する DAF の実験を行なった.

### 実験方法

実験参加者 (話者) は正常聴力を持つ 22–26 歳の日本人男性 4 名であった. 実験は防音室内で行なわれた. 話者の音声はマイクロフォン (Sennheiser MHK 416 P48) からオーディオインタフェース (Roland QUAD-CAPTURE) を経由して, PC (LG Sharkoon, with OS Ubuntu Studio 14.04 Linux) により収録された. サンプリング周波数は 44100 Hz, 量子化ビット数は 16 bit とした. 音声信号は, 実験条件に応じた遅延が挿入され, インタフェースを經由してアンプ (audio-technica AT-HA5000) で増幅された後, ヘッドフォン (Sennheiser HD280) または骨導トランスデューサ (TEAC HP-F200) によって話者に呈示された. 実験機構全体での潜時は 13.2 ms であった. DAF の実験機構の概略図を図 5.1 に示す.

話者が発音する実験刺激として, 日本語の 3 つの 4 モーラ単語 (*/sjabusjabu/*, */takabisja/*, */yuuwaku/*) を用いた. 話者は無遅延条件の下で各単語を 5 回ずつ発音した後, 6 つの遅延条件 (25, 50, 100, 150, 200, 400 ms) の下で発話した. 遅延条件下では, 話者が 1 単語を 30 回発音する間, 1 回ずつ遅延条件がランダムに変更され, 1 条件あたり各単語が 5 回ずつ発音されるようにした. この無遅延条件–遅延条件のセットを, 気導呈示・骨導呈示によってそれぞれ 1 セットずつ行なった. 骨導呈示時には話者に耳栓を装着させ, 両呈示時ともに刺激音声以外の音や雑音を呈示しなかった. 刺激音声の音圧レベルは始めに話者によって任意に調節され, 実験中の変更は為されなかった.

遅延音声が発話スタイルに与える影響を定量化するため, 無遅延条件時に対する各遅延条件時の発話時間比 ( $R_d$ ) と, 各遅延条件時の非流暢箇所の総数 ( $N_d$ ) を調査した.  $R_d$

は (5.1) 式および (5.2) 式に従って求められた。

$$R_d = \frac{1}{N_w} \sum_{i=1}^{N_w} \frac{T_{d,i}}{\bar{T}} \quad (5.1)$$

$$\bar{T} = \frac{1}{N_w} \sum_{i=1}^{N_w} T_i \quad (5.2)$$

ここで、 $T_i$ 、 $T_{d,i}$  はそれぞれ無遅延条件および  $d$  [ms] の遅延条件 ( $d = 25, 50, \dots, 400$ ) における  $i$  試行目の発話時間を、 $N_w$  は 1 人の話者の各遅延条件における 1 単語の試行数を表す ( $N_w = 5$ ,  $i = 1, 2, \dots, N_w$ )。収録された音声波形に対応する  $T_i$ 、 $T_{d,i}$  の例を図 5.2 に示す。さらに、先行研究 [52] と同様にして、単語発音中の部分延長と部分繰り返しを非流畅箇所と定義し、(5.3) 式に従って  $N_d$  を求めた。

$$N_d = \sum_{j=1}^{N_c} s_{d,j} \quad (5.3)$$

ここで、 $N_c$  は各遅延条件における、全話者・全単語での試行総数を表す ( $N_c = 60$ ,  $j = 1, 2, \dots, N_c$ )。また  $s_{d,j}$  は、 $d$  [ms] の遅延条件 ( $d = 25, 50, \dots, 400$ ) における  $j$  試行目の発音に非流畅箇所が見られた場合に  $s_{d,j} = 1$  を、見られなかった場合に  $s_{d,j} = 0$  と定義されるものとする。図 5.2 に、 $s_{d,j} = 0$  と  $s_{d,j} = 1$  となる音声波形の例が示されている。

## 結果

遅延条件による  $R_d$  の推移について、結果の一例を図 5.3 に示す。図 5.3(a) および図 5.3(b) では、気導呈示時・骨導呈示時の両方で、遅延条件に依存した  $R_d$  の増加が見られた。気導呈示時には、遅延の増加とともに  $R_d$  が徐々に増加し、100–200 ms 付近でピークが見られた。一方、骨導呈示時には 25 ms の遅延条件時まで急激に  $R_d$  が増加する傾向が見られ、それ以降は大きな変化が見られなかった。 $R_d$  曲線の概形は、発音された単語によりやや異なった。特に、/sjabusjabu/ 発音時の標準偏差は他の 2 つの単語発音時に比べ大きい傾向があった。また、 $R_d$  曲線の概形は参加者によって大きく異なっており、話者 C (図 5.3(c)) では遅延条件によらず  $R_d$  の大きな変化は見られなかった。

遅延条件による  $N_d$  の推移を図 5.4 に示す。気導呈示時・骨導呈示時の両方で、遅延条件に依存した  $N_d$  の増加が見られた。両呈示時ともに、200 ms 以下の遅延条件において  $N_d$  が大きい傾向が見られた。全体として、 $N_d$  は骨導呈示時のほうが気導呈示時に比べてはるかに大きかった。/sjabusjabu/ 発音時の非流畅試行の数は、他の 2 単語発音時に比べはるかに大きかった。また、非流畅試行の数は、被験者により大きく異なっていた (A: 13, B: 0, C: 7, D: 20)。

## 考察とまとめ

気導・骨導音声を用いた DAF の実験結果から、下記のことことが分かった。

- 発話時間の増加や吃りなどの発話スタイル変化は，気導呈示時だけでなく骨導呈示時にも見られた．
- 発話スタイル変化の傾向は，発音する単語や話者によって異なり，気導呈示時と骨導呈示時の間での一定した傾向の違いは発見されなかった．
- 全体として，発話スタイルの変化は骨導呈示時でのほうが気導呈示時に比べ大きい傾向が見られた．

この結果から，発話に対して気導・骨導の両方の聴覚フィードバックが関係していることが示唆される．骨導呈示時での大きな発話スタイルの変化から，気導に比べ骨導の聴覚フィードバックによる知覚-生成間の時間的整合性の破綻は，気導の聴覚フィードバックよりも発話に大きな影響を与えたと考えられる．また，上記の結果は，気導・骨導での聴覚フィードバックの音響的特徴の伝達速度が異なることを示している可能性がある．

今後は，発音する単語の種類や刺激数，参加者数を再考の上，さらに深い検討をする必要があると考えられる．

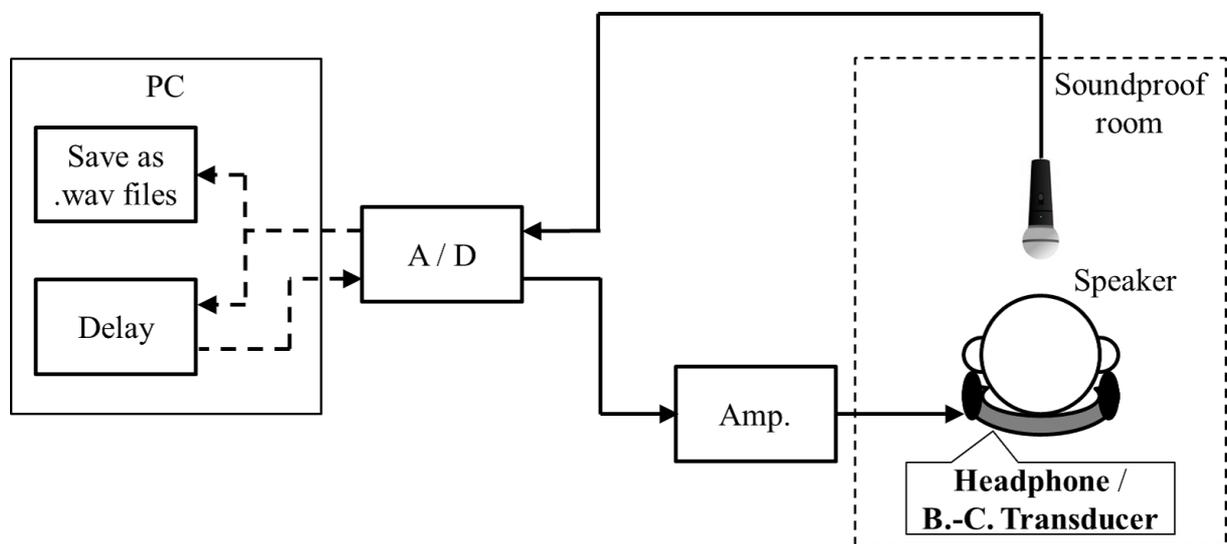


図 5.1: DAF の実験機構の概略

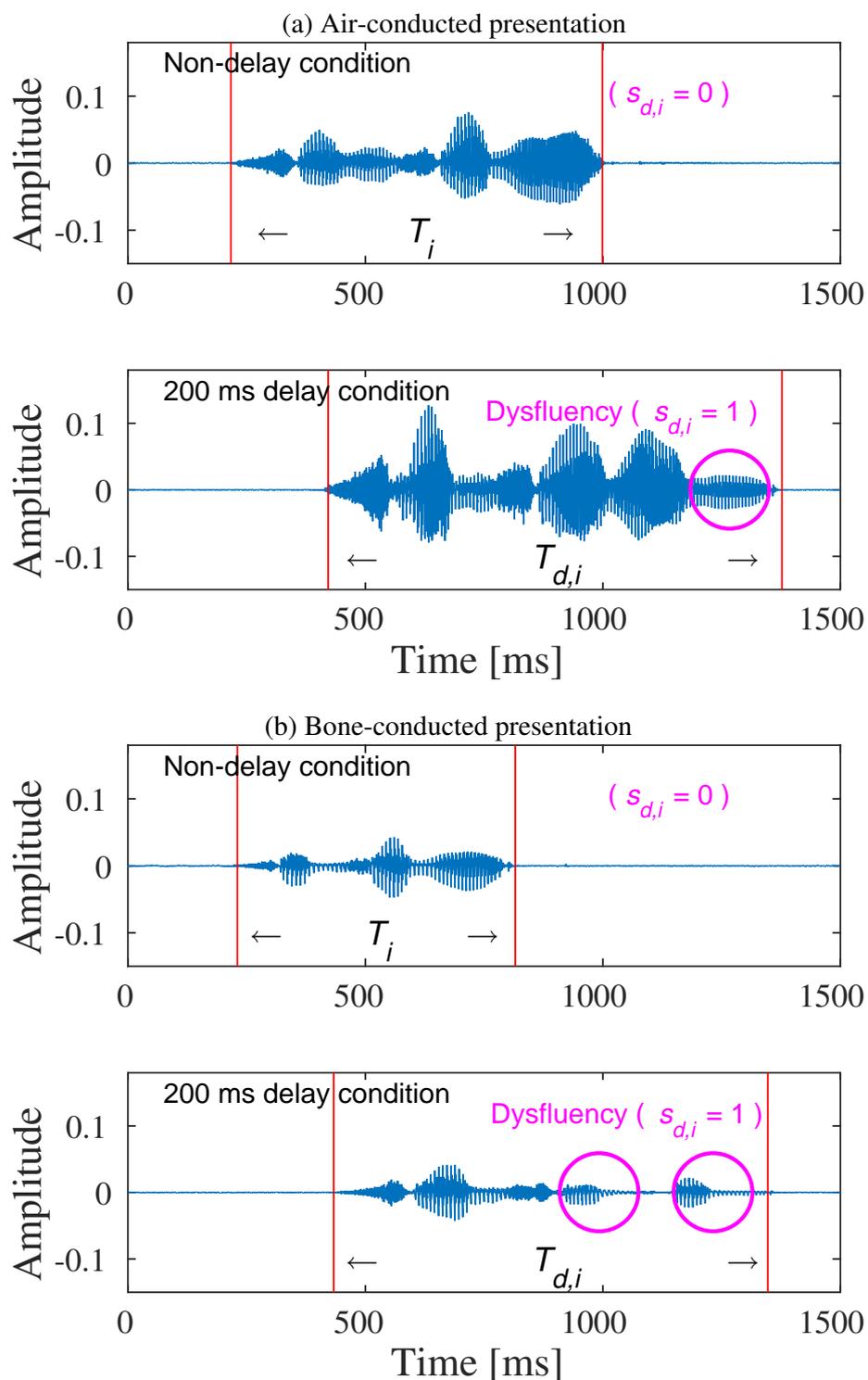


図 5.2: 収録された音声波形中の  $T_i$ ,  $T_{d,i}$  ( $d = 200$ ) および  $s_{d,i}$  の例 (話者 A による /sjabusjabu/ 発声時の (a) 気導呈示時, (b) 骨導呈示時)

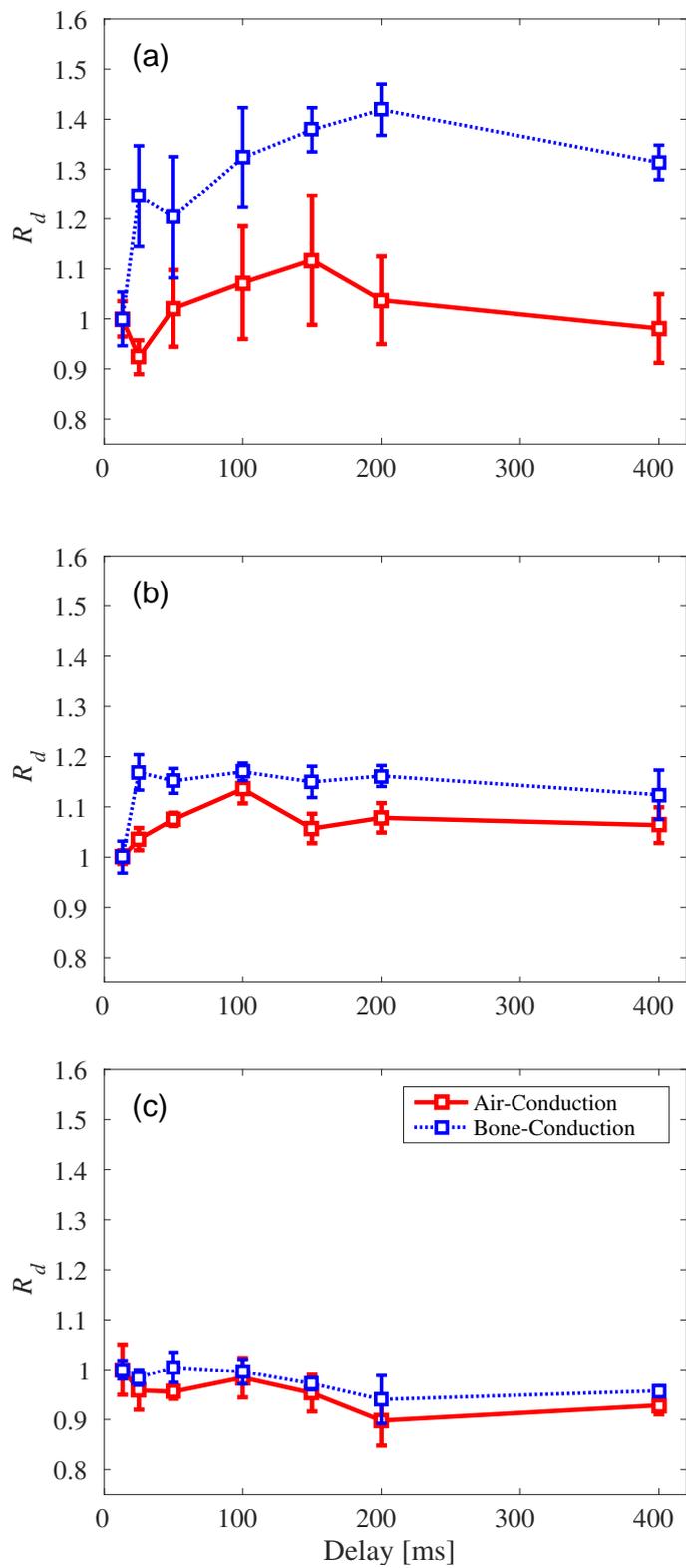


図 5.3: 遅延条件による  $R_d$  の推移 : (a) 話者 A による /sjabusjabu/ 発音時, (b) 話者 B による /takabisja/ 発音時, (c) 話者 C による /yuuwaku/ 発音時 (エラーバー: 標準偏差)

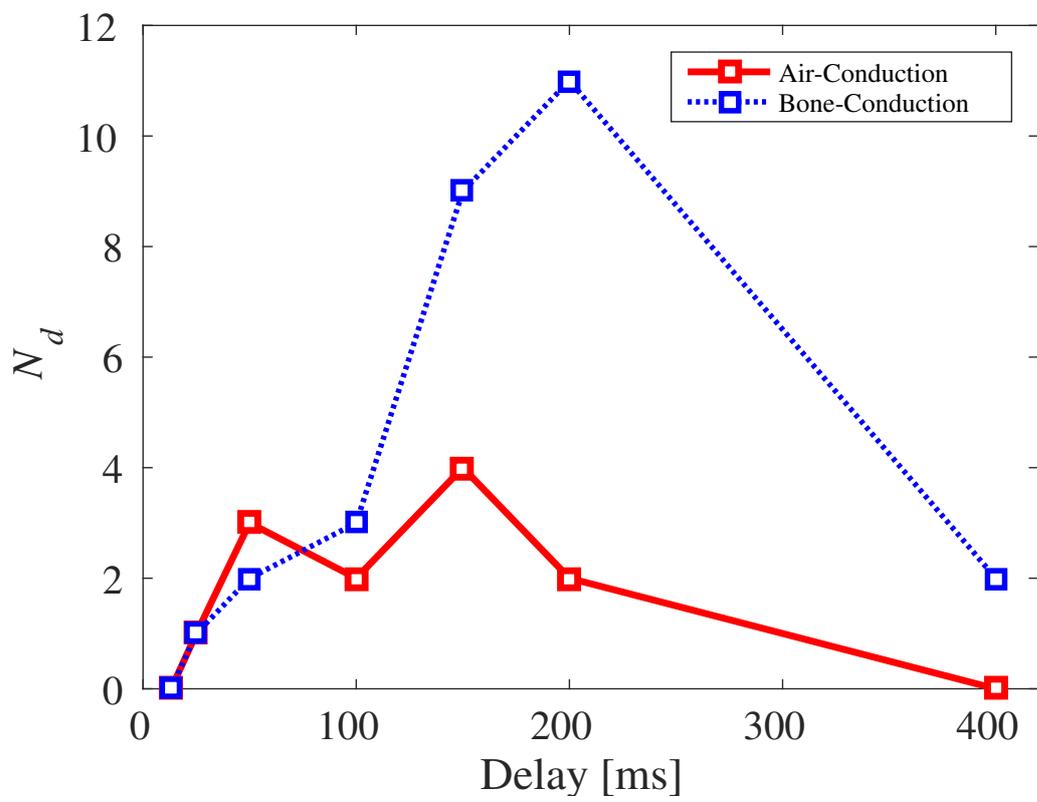


図 5.4: 遅延条件による  $N_d$  の推移

## 参考文献

- [1] 片桐功, 吉川文, 岸啓子, 久保田慶一, 長野俊樹, 白石美雪, 高橋美都, 三浦裕子, 茂手木潔子, 塚原康子, 榎崎洋子, “はじめての音楽史 増補改訂版,” 音楽之友社, 東京, 2009.
- [2] R. Aiello and J. A. Sloboda, “Musical Perceptions,” Oxford University Press, Oxford, 1994. (大串健吾 訳, “音楽の認知心理学,” 誠信書房, 東京, 1998.)
- [3] 剣持秀紀, “歌声合成システム VOCALOID の開発,” システム/制御/情報, Vol. 56, No.5, pp. 244–248, 2012.
- [4] 重野純, “聴覚・ことば,” 新曜社, 東京, 2006.
- [5] 加我君孝, “聴覚と音痴,” 音声言語医学, Vol. 41, pp. 266–271, 1999.
- [6] S. D. Bella, M. Berkowska and J. Sowinski, “Disorders of pitch production in tone deafness,” *Frontiers in Psychology*, Vol. 2, No.164, pp. 1–11, 2011.
- [7] 米沢義道, 伊東一典, 平野圭蔵, “音痴の音高感覚に関する検討,” 電子情報通信学会論文誌, Vol. J71-A, No.8, pp. 1532–1538, 1988.
- [8] 重野純, “絶対音感と音痴,” 音声言語医学, Vol. 41, pp. 260–265, 2000.
- [9] T. Yuba, “Method for Curing Motor-Related Off-Key Singing,” 音声言語医学, Vol. 41, pp. 250–254, 1999.
- [10] P. Q. Pfordresher and S. Brown, “Poor-Pitch Singing in the Absence of “Tone Deafness”,” *Music Perception*, Vol. 25, No.2, pp. 95–115, 2007.
- [11] D. A. Nikjeh, J. J. Lister and S. A. Frisch, “The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians,” *J. Acoust. Soc. Am.*, Vol. 125, No.1, pp. 328–338, 1986.
- [12] E. Bradshaw and M. A. McHenry, “Pitch Discrimination and Pitch Matching Abilities of Adults who Sing Inaccurately,” *J. Voice*, Vol. 19, No.3, pp. 431–439, 2005.
- [13] M. Unoki and K. Nishimoto, “Improvements to creativity in singing abilities based on perspective of studies on interaction between speech production and auditory perception,” *Proc. KICSS2012*, pp. 157–160, 2012.

- [14] 平井重行, 片寄晴弘, 井口征士, “歌の調子外れに対する治療支援システム,” 電子情報通信学会論文誌, Vol. J84-D-II, No. 9, pp. 1933–1941, 2001.
- [15] 村尾忠廣, “視覚フィードバックによる音痴の治療,” 音声言語医学, Vol. 41, No.3, pp. 255–259, 2000.
- [16] P. Denes and E. Pinson, “The Speech Chain, 2nd Ed.,” W. H. Freeman and Co., 1993.
- [17] A. Postma, “Detection of errors during speech production: a review of speech monitoring models,” *Cognition*, Vol. 77, pp. 97–131, 2000.
- [18] 河原英紀, “聴覚フィードバックの発声への影響,” 日本音響学会誌, Vol. 59, No. 11, pp. 670–675, 2000.
- [19] P. Madaule, “Listening and Singing,” *NATS Journal of Singing*, 2001.
- [20] J. Sundberg, “The Science of the Singing Voice,” Northern Illinois University Press, Illinois, 1987.
- [21] 古井貞熙, 音声情報処理, 森北出版, 東京, 1998.
- [22] 日本音響学会, 新版 音響用語辞典, コロナ社, 東京, 2003.
- [23] R. D. Kent and C. Read, “The Acoustic Analysis of Speech,” Singular Publishing Group, Inc., California, 1992. (荒井隆行, 菅原勉 訳, “音声の音響分析,” 海文堂, 東京, 1996.)
- [24] R. F. Port and J. Dalby, “Consonant/Vowel ratio as a cue for voicing in English,” *Perception & Psychophysics*, Vol. 2, pp. 141–152, 1982.
- [25] G. E. Peterson and H. L. Barney, “Control Methods Used in a Study of the Vowels,” *J. Acoust. Soc. Am.*, Vol. 24, No.2, pp. 175–184, 1952.
- [26] M. L. Ng, H. R. Gilbert and J. W. Lerman, “Fundamental Frequency, Intensity, and Vowel Duration Characteristics Related to Perception of Cantonese Alaryngeal Speech,” *Folia Phoniatrica et Logopaedica*, Vol. 53, pp. 36–47, 2001.
- [27] A. Paeschke, M. Kienast and W. F. Sendlmeier, “ $F_0$ -Contours in Emotional Speech,” *Proc. ICPHS99*, pp. 929–932, 1999.
- [28] S. Rosen and P. Howell, “Signals and Systems for Speech and Hearing,” Academic Press, London, 1991.

- [29] G. E. Peterson and H. L. Barney, “Control Methods Used in a Study of the Vowels,” *J. Acoust. Soc. Am.*, Vol. 24, No. 2, pp. 175–184, 1952.
- [30] W.-H. Tsai, H.-C. Lee, “Automatic Evaluation of Karaoke Singing Based on Pitch, Volume, and Rhythm Features,” *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 20, No. 4, pp. 1233–1243, 2012.
- [31] D. Deutsch, “The Psychology of Music,” Academic Press, Waltham, 1982. (寺西立年, 大串健吾, 宮崎謙一 訳, “音楽の心理学 (上),” 西村書店, 新潟, 1987.)
- [32] J. Sundberg, “Articulatory differences between spoken and sung vowels in singers,” *STL-QPSR*, Vol. 10, No. 1, pp. 33–46, 1969.
- [33] S. R. Livingstone, K. Pech and F. A. Russo, “Acoustic differences in the speaking and singing voice,” *Proc. Meet. on Acoust.*, Vol. 19, pp. 1–5, 2013.
- [34] T. Nakano and M. Goto, “VocaListener: A Singing-to-Singing Synthesis System Based on Iterative Parameter Estimation,” *Proc. SMC 2009*, pp. 343–348, 2004.
- [35] 大石康智, 後藤真孝, 伊藤克巨, 武田一哉, “スペクトル包絡と基本周波数の時間変化を利用した歌声と朗読音声の識別,” *情報処理学会論文誌*, Vol. 47, No. 6, pp. 1822–1830, 2006.
- [36] R. J. Zatorre and S. R. Baum, “Musical Melody and Speech Intonation: Singing a Different Tune?,” *PLoS Biology*, Vol. 10, No. 7, pp. 1–6, 2012.
- [37] 剣持秀紀, “歌声合成技術の動向,” *日本音響学会誌*, Vol. 67, No. 1, pp. 46–50, 2011.
- [38] H. Fujihara, M. Goto, T. Kitahara and H. G. Okuno, “A Modeling of Singing Voice Robust to Accompaniment Sounds and Its Application to Singer Identification and Vocal-Timbre-Similarity-Based Music Information Retrieval,” *Proc. IEEE Trans. Aud. Spee. & Lang. Proc.*, Vol. 10, No. 7, pp. 1–6, 2012.
- [39] H. Mori, W. Odagiri and H. Kasuya, “ $F_0$  Dynamics in Singing: Evidence from the Data of a Baritone Singer,” *IEICE Trans. Inf. & Syst.*, Vol. 87-D, No. 5, pp. 1086–1092, 2004.
- [40] D. Gerhard, “Pitch-Based Acoustic Feature Analysis for the Discrimination of Speech and Monophonic Singing,” *Canadian Acoustics*, Vol. 130, No. 3, pp. 152–153, 2002.
- [41] 齋藤毅, 辻直哉, 鶴木祐史, 赤木正人, “歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析,” *日本音響学会誌*, Vol. 64, No. 5, pp. 267–277, 2008.

- [42] K. R. Scherer, J. Sundberg, L. Tamarit and G. L. Salomão, “Comparing the acoustic expression of emotion in the speaking and the singing voice,” *Computer Speech and Language*, Vol. 29, pp. 218–235, 2013.
- [43] T. Saitou, M. Goto, M. Unoki and M. Akagi, “Speech-to-Singing Synthesis: Converting Speaking Voices to Singing Voices by Controlling Acoustic Features Unique to Singing Voices,” *Proc. WASPAA 2007*, pp. 215–218, 2007.
- [44] S. Wang, “Singer’s high formant associated with different larynx position in styles of singing,” *J. Acoust. Soc. Jpn.(E)*, Vol. 7, No. 6, pp. 303–314, 1986.
- [45] J. Sundberg, “A perceptual function of the ‘singing formant’,” *STL-QPSR*, Vol. 13, No. 2-3, pp. 61–63, 1972.
- [46] E. Lombard, “Le signe de l’elevation de la voix,” *Annales des Maladies de l’Oreille, du Larynx, du Nez et du Pharynx*, Vol. 37, No. 2, pp. 101–119, 1911. (translated by P. H. Mason, “The sign of the elevation of the voice,” 2006.)
- [47] 日本音響学会, “音響キーワードブック,” コロナ社, 東京, 2016.
- [48] I. Ipsic, “Speech Technologies,” Intech, Winchester, 2011.
- [49] J.-C. Junqua, A. Fincke and K. Field, “The Lombard Effect: A Reflex to Better Communicate with Others in Noise,” *Proc. ICASSP 1999*, Vol. 4, pp. 2083–2086, 1999.
- [50] B. S. Lee, “Effects of Delayed Speech Feedback,” *J. Acoust. Soc. Am.*, Vol. 22, No. 6, pp. 824–826, 1950.
- [51] S. Okazaki, K. Mori and C. Cai, “Effects of delayed auditory feedback on the vocal time-reproduction,” *Acoust. Sci. & Tech.*, Vol. 31, No. 6, pp. 408–409, 2010.
- [52] A. Stuart, J. Kalinowski, M. P. Rastatter and K. Lynch, “Effect of delayed auditory feedback on normal speakers at two speech rates,” *J. Acoust. Soc. Am.*, Vol. 111, No. 5, pp. 2237–2241, 2002.
- [53] 河原英紀, “声を使って聴覚を探る,” *日本音響学会誌*, Vol. 53, No. 9, pp. 731–737, 1997.
- [54] J. V. Borsel, G. Reunes and N. V. den Bergh, “Delayed auditory feedback in the treatment of stuttering: clients as consumer,” *Int. J. Lang. Comn. Dis.*, Vol. 38, No. 2, pp. 119–129, 2003.

- [55] H. Kawahara, H. Kato and J. C. Williams, “Effects of Auditory Feedback on  $F_0$  Trajectory Generation,” *Spoken Language*, Vol. 1, pp. 287–290, 1996.
- [56] T. A. Burnett, M. B. Freedland, C. R. Larson and T. C. Hain, “Voice  $F_0$  responses to manipulations in pitch feedback,” *J. Acoust. Soc. Am.*, Vol. 103, No. 6, pp. 3153–3161, 1998.
- [57] C. R. Larson, T. A. Burnett and Swathi Kiran, “Effects of pitch-shift velocity on voice  $F_0$  responses,” *J. Acoust. Soc. Am.*, Vol. 107, No. 1, pp. 559–564, 2000.
- [58] Z. Chen, P. Liu, J. A. Jones, D. Huang and H. Liu, “Sex-related differences in vocal responses to pitch feedback perturbations during sustained vocalization,” *J. Acoust. Soc. Am.*, Vol. 128, No. 6, pp. 355–360, 2010.
- [59] H. Liu, N. M. Russo and C. R. Larson, “Age-related differences in vocal responses to pitch feedback perturbations: A preliminary study,” *J. Acoust. Soc. Am.*, Vol. 127, No. 2, pp. 1042–1046, 2010.
- [60] C. R. Larson, J. Sun and T. C. Hain, “Effects of simultaneous perturbations of voice pitch and loudness feedback on voice  $F_0$  and amplitude control,” *J. Acoust. Soc. Am.*, Vol. 121, No. 5, pp. 2862–2872, 2007.
- [61] A. Toyomura, S. Koyama, T. Miyamoto, A. Terao, T. Omori, H. Murohashi and S. Kuriki, “Neural Correlates of Auditory Feedback Control in Human,” *Neuroscience.*, Vol. 146, pp. 499–503, 2007.
- [62] R. J. Zatorre, A. C. Evans, E. Mayer and A. Gjedde, “Lateralization of phonetic and pitch discrimination in speech processing,” *Science*, Vol. 256, pp. 846–849, 1992.
- [63] R. J. Zatorre and S. Samson, “Role of the right temporal neocortex in retention of pitch in auditory short-term memory,” *Brain A Journal of Neurology*, Vol. 114, pp. 2403–2417, 1991.
- [64] T. M. Donath, U. Natke and K. Th. Kalveram, “Effects of frequency-shifted auditory feedback on voice  $F_0$  contours in syllables,” *J. Acoust. Soc. Am.*, Vol. 111, No. 1, pp. 357–366, 2002.
- [65] S. H. Chen, H. Liu, Y. Xu and C. R. Larson, “Voice  $F_0$  responses to pitch-shifted voice feedback during English speech,” *J. Acoust. Soc. Am.*, Vol. 121, No. 2, pp. 1157–1163, 2007.
- [66] D. W. Purcell and K. G. Munhall, “Compensation following real-time manipulation of formants in isolated vowels,” *J. Acoust. Soc. Am.*, Vol. 119, No. 4, pp. 2288–2297, 2006.

- [67] S. Cai, S. S. Ghosh, F. H. Guenther and J. S. Perkell, “Focal Manipulations of Formant Trajectories Reveal a Role of Auditory Feedback in the Online Control of Both Within-Syllable and Between-Syllable Speech Timing,” *J. Neuroscience.*, Vol. 31, No. 45, pp. 16483–16490, 2011.
- [68] 松岡理絵, X. Lu, 党建武, 赤木正人, “調音情報を考慮した聴覚系と発話系の相互作用に関する検討,” 日本音響学会聴覚研究会資料, Vol. 34, pp.1–6, 2003.
- [69] D. A. Nikjeh, J. J. Lister and S. A. Frisch, “The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians,” *J. Acoust. Soc. Am.*, Vol. 125, No. 1, pp. 328–338, 2009.
- [70] Y. Lévêque, A. Giovanni and D. Schön, “Pitch-Matching in Poor Singers: Human Model Advantage,” *J. Voice*, Vol. 26, No. 3, pp. 293–298, 2009.
- [71] D. Keough and J. A. Jones, “The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing,” *J. Acoust. Soc. Am.*, Vol. 126, No. 2, pp. 837–846, 2009.
- [72] S. Iijima, S. Seikoba, S. Ishimitsu and M. Nakayama, “Effects of High-Pass Filtered and Pitch-Shifted Auditory Feedback on a Singer’s Pitch,” *IEIC Express Letters Part B: Applications*, Vol. 6, No. 12, pp. 3285–3289, 2015.
- [73] C. Leydon, J. J. Bauer and C. R. Larson, “The role of auditory feedback in sustaining vocal vibrato,” *J. Acoust. Soc. Am.*, Vol. 114, No. 3, pp. 1575–1581, 2003.
- [74] D. Mürbe, F. Pabst, G. Hofmann and J. Sundberg, “Significance of Auditory and Kinesthetic Feedback to Singer’s Pitch Control,” *J. Voice*, Vol. 16, No. 1, pp. 44–51, 2002.
- [75] C. Watts, J. Murphy and K. B.-Burroughs, “Pitch Matching Accuracy of Trained Singers, Untrained Subjects with Talented Singing Voices, and Untrained Subjects with Nontalented Singing Voices in Conditions of Varying Feedback,” *J. Voice*, Vol. 17, No. 2, pp. 185–194, 2003.
- [76] D. J. Levitin, “Absolute memory for musical pitch: Evidence from the production of learned melodies,” *Perception & Psychophysics*, Vol. 56, No. 4, pp. 414–423, 1994.
- [77] R. J. Zatorre, “Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function,” *Nature Neuroscience*, Vol. 6, No. 7, pp. 692–695, 2003.

- [78] 宮崎謙一, “「絶対音感」はどこまで分かったか?,” 日本音響学会誌, Vol. 60, No. 11, pp. 682–688, 2004.
- [79] S. Stenfelt, “Acoustic and physiological aspects of bone conduction hearing,” *Advances in Oto-Rhino-Laryngology*, Vol. 71, pp. 10–21, 2011.
- [80] S. Stenfelt, T. Wild, N. Hato and R. L. Goode, “Factors contributing to bone conduction: The outer ear,” *J. Acoust. Soc. Am.*, Vol. 113, No. 2, pp. 902–913, 2003.
- [81] S. Stenfelt and R. L. Goode, “Bone-Conducted Sound: Physiological and Clinical Aspects,” *Otology & Neurotology*, Vol. 26, No. 6, pp. 1245–1261, 2005.
- [82] J. Tonndorf, “Compressional Bone Conduction in Cochlear Models,” *J. Acoust. Soc. Am.*, Vol. 34, No. 8, pp. 1127, 1962.
- [83] S. Y. Won and J. Berger, “Estimating transfer function from air to bone conduction using singing voice,” *Proc. International Computer Music Conference*, 2005.
- [84] T. Tamiya and T. Shimamura, “Reconstruction Filter Design for Bone-Conducted Speech,” *Proc. Interspeech 2004*, 2004.
- [85] K. Kondo, T. Fujita and K. Nakagawa, “On Equalization of Bone Conducted Speech for Improved Speech Quality,” *Proc. ISSPPT 2006*, pp. 426–431, 2006.
- [86] P. Tran, T. Letowski and M. McBride, “Bone Conduction Microphone: Head Sensitivity Mapping for Speech Intelligibility and Sound Quality,” *Proc. ICALIP 2008*, pp. 107–111, 2008.
- [87] S. Reinfeldt, P. Östli, B. Håkansson and S. Stenfelt, “Hearing one’s own voice during phoneme vocalization-Transmission by air and bone conduction,” *J. Acoust. Soc. Am.*, Vol. 128, No. 2, pp. 751–762, 2010.
- [88] X. Wang, K. Honda, J. Dang, and J. Wei, “Vocal Responses to Frequency Modulated Composite Sinewaves via Auditory and Vibrotactile Pathways,” *Proc. ICASSP2015*, pp. 4355–4359, 2015.
- [89] 石田舞, “調波複合音のスペクトル包絡に応じて変化する音の高さに関する研究,” 北陸先端科学技術大学院大学 情報科学研究科 修士論文, 2010.
- [90] 山本克彦, “骨導音知覚特性に関する調査研究,” 北陸先端科学技術大学院大学 情報科学研究科 課題研究報告書, 2015.
- [91] G. von Békésy, “Zur Theorie des Hörens bei der Schallaufnahme durch Knochenleitung,” *Ann. Phys.*, Vol. 405, No.1, pp. 111–136, 1932.

# 謝辞

本課題研究の遂行に当たり，多大な助言と懇切丁寧かつ，情熱溢れる御指導をして頂きました鶴木祐史 准教授に心から感謝致します。また，研究室会議をはじめ，様々な機会  
で助言をして頂きました赤木正人 教授に心から感謝致します。さらに，日ごろの研究討  
論や実験に関する多くの御指導，助言をして頂きました宮内良太 助教，木谷俊介 助教，  
森川大輔 助教，久保理恵子 博士に心から感謝致します。加えて，夏期合同合宿等で，本  
調査研究に関する貴重なご指摘，助言を頂きました党建武 教授，田中宏和 准教授，末光  
厚夫 助教に深くお礼申し上げます。研究活動に関してだけでなく，公私にわたりお世話  
になりました赤木・鶴木研究室の諸氏に改めて深く御礼申し上げます。

# 研究業績リスト

## 本研究に関連する研究業績

### 論文

1. T. Toya, D. Ishikawa, R. Miyauchi, K. Nishimoto and M. Unoki, “Study on Effects of Speech Production during Delayed Auditory Feedback for Air-Conducted and Bone-Conducted Speech,” *Journal of Signal Processing*, Vol. 20, No. 4, 2016.

### 国際会議

1. T. Toya, D. Ishikawa, R. Miyauchi, K. Nishimoto and M. Unoki, “Study on Effects of Speech Production during Delayed Auditory Feedback for Air-Conducted and Bone-Conducted Speech,” 2016 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP'16), Honolulu, Mar. 2016.

## その他の研究業績

### 口頭発表

1. 鳥谷輝樹, 宮内良太, 鷓木祐史, “心理物理的同調曲線による抑圧性マスキングの検討,” 平成 27 年度電気関係学会北陸支部連合大会, 石川, 2015 年 9 月.

### ポスター発表

1. 鳥谷輝樹, 宮内良太, 鷓木祐史, “心理物理的同調曲線による抑圧性マスキングの検討,” 第 7 回北陸合同音響セミナー ビギナー成果発表会, 石川, 2015 年 12 月.
2. 石川大介, 鳥谷輝樹, 鷓木祐史, “Puredata を用いた遅延聴覚フィードバックシステムの構築,” 第 7 回北陸合同音響セミナー ビギナー成果発表会, 石川, 2015 年 12 月.