

音声科学におけるデータ収集法：韻律研究を中心にして (音声研究法管見)

前川喜久雄

(国立国語研究所 言語行動研究部 第2研究室)

筆者は言語学の教育をうけ、現在は音声研究をに携わっている。ここ10年ほどのあいだは日本語の韻律現象を実験的な手法で研究してきた。気の散りやすい性格のせいで、いろいろな研究を並行して進めてきているが、そのために、種々の音声データの取得と解析を体験することはできた。

以下、本稿の前半では音声データの取得における具体的な問題点を体験に基づきながら思いつくままに列挙する。つづいて後半では、実際の研究事例をとりあげて音声研究の実際について紹介を試みる。データをとるという作業は研究の目的と切り離しては議論できないからである。

I 音声科学におけるデータの種類・性質・問題点

1 音声信号

音声信号はテープレコーダの発明以来簡単に記録できるようになった。現在では種々のデジタル録音も可能であり、DAT(digital audio tape)が標準的な記録装置となっている。最近普及が著しいMDは記憶容量をあげるために周波数圧縮をおこなっているので利用しないほうが安全である。DATで収録したデータは多くの場合、計算機を利用して分析することになる、デジタル信号のまま計算機にとりこむこともできるが、そのためには特殊なインターフェースが必要になり、現状では価格も安くはない。そこで、DATの信号を一旦アナログ信号として出力し、アナログ・デジタル変換ボードを経由して計算機にとりこむことがある。そこで問題になるのがA/Dボードの性能である。計算機のディスクに直接音声信号を書き込む場合も同様である。多くの製品を比較検討することは困難であるが、おおまかな目安として、ノート型パソコンのA/Dボードはデスクトップ型にくらべて性能が劣り、内部雑音も大きい。デスクトップ型の場合も、A/D用チップをマザーボードに内蔵したタイプよりも、拡張スロットに独立したボードをさしこむタイプの方が安心できるだろう。

A/D変換ではサンプリング周波数とサンプリング精度を指定しなければならない。後者は通常16ビットにしておけばよいが、前者は分析の目的を考えて指定する必要がある。デジタル信号処理では実際に分析可能な周波数帯域はサンプリング周波数の1/2までである。サンプリング周波数が低すぎると、高い周波数領域の現象を観察することができない。反対にサンプリング周波数をむやみに高く設定するのも問題である。ディスク資源を無駄に消費し、信号処理に時間がかかり、分析結果を表示した際にも周波数軸が細かすぎて観察に支障をきたすことになりかねない。分析の目的をよく考える必要がある。

音声信号の記録にはマイクが必要である。マイクの性能と使用法はデータの質に決定的な影響をおよぼす。マイクには大別してダイナミック型と(エレクトリック)コンデンサ型があるが、多くの場合、後者を使用しておけばよい。複数話者の対話を収録する場合は、話者ひとりにマイク1個を用意し、テープレコーダの別チャンネルに記録するのが原則である。3名以上の対話を高精度で収録する場合は、通常の2チャンネルのテープレコーダを複数用意するか、多チャンネルのデータレコーダを利用する(今回のセミナーにおける小磯のハンドアウト参照)。

音声信号には、波形、振幅、位相の情報が含まれているが、このうち振幅情報に注目した分析をおこなうためには、マイクと唇の距離を一定にたもっておく必要がある。厳密な制御は不可能であるが、ヘッドセット型マイクを利用することによって問題を改善することはできる。

また位相情報に依存した分析手法を適用する場合にも独自の注意が必要になる。位相のずれや反転は耳で音を聞いてもほとんど感知することができないが、零交差を利用したピッチ抽出装置や、波形からピッチ周期を決定した上でピッチ同期のスペクトル分析をおこなうソフトウェアなどは、位相の変化による影響をこらむる。注意点として、第一にダイナミック型マイクは位相がずれやすいので、コンデンサ型を用いるべきである。第二に、コンデンサ型マイクを用いても、信号が種々のアンプを経由してゆく過程で位相が上下反転してしまうことがある。マイクからテープに至る経路に何個のアンプが介在しているかを知っておくとよい。なお、位相の上下反転だけならば、計算機上の簡単な処理でもとの信号に復元させることができる。

最後に音声を収録する環境の問題がある。フィールドで音声を収録する場合、種々の環境ノイズを避け難いことが多いが、ノイズの存在は意識的に注意していないと気づかないものである。時計の音、風の音、車の騒音、マイクをおいた机の振動、タイピンマイクと衣服のこすれ、等々。また静かな室内であっても、反響の強い場所は望ましくない。音声の実験的研究を恒常的におこなうならば簡単な防音室を設置しておいた方がよい。また、分析の目的によっては部屋の大きさによって決まる定在波の影響をさけるために無響室内での収録が必要となることもある。

収録された音声データは通常、音響的な分析にかけられることが多いだろう。音響分析について指摘すべき注意点は山ほどあるがすべて省略せざるをえない。ひとつだけよく受ける質問に答えておこう。それは音響分析のためのソフトウェアとして何がよいかという質問である。現在では種々のソフトウェアが有償/無償で公開されている。プラットフォームをDOS/V & Windows だけに限って一部を紹介すると以下のようである（日本音響学会誌 55-3 に国産製品の関連記事がある）。

名称	開発/取り扱い	備考	価格(円)
SP4WIN	NTT アドバンスドテクノロジー Email sp4win@kana.ntt-at.co.jp	Windows で稼動。上位バージョンはユーザー関数定義可能。	58,000~98,000
Acoustic Core	アルカディア URL http://www.arcadia.co.jp	Windows で稼動 ピッチ抽出・フォルマント抽出機能なし	98,000 アカデミック価格あり
ISL	Kay Elemetrics (USA) / リオン	DOS で稼動。オプションで LPC 合成機能。 Windows 版も開発は終了。 国内販売予定は?	オプションによる。基本セットが 100 万前後だったと思う
Multi Speech	Kay Elemetrics (USA) / リオン	Windows で稼動 IPA によるラベリング	20 万程度
L★Voice	宇都宮大学粕谷研究室 / リバティシステム rinji@libertysystem.co.jp	オプションで ARX 法による音源パラメータ推定をサポート	オプションによる。フルセットで 15 万程度。
Praat	Univ. Amsterdam http://fonsg3.let.uva.nl/praat/praat.html	Windows の他各種 Unix で稼動。	シェアウェア
EMU	Macquarie Univ. http://www.shlrc.mq.edu.au/emu/	音声データベース検索と階層ラベリング機能。	フリーウェア 近く書籍の付録としても販売

これらに限っていえば、基本的な音響分析（持続時間測定、ピッチ測定、スペクトログラム作成、フォルマント抽出など）については一応同じ水準の分析が可能である。もちろん、操作性はそれぞれに異なるし、同じピッチ抽出でも、あるソフトウェアは零交差分析、別のソフトは自己相関法といった分析アルゴリズム

の違いがあり、その結果、ソフトウェアごとに分析の得手不得手があることは確かである。さらにサポートしている機能も全体を見れば大幅に相違しており、Praat や EMU は大学で開発された本格的・包括的な音声研究用ツール群である。

しかし、音響分析を実施するにあたって、何よりも大きな問題はユーザーの側にある。音響分析ソフトウェアはワープロなどとは違ってあくまで研究用ツールである。したがって研究者として最低限必要な基礎知識なしに操作しても、その結果の正しさは誰も保証してくれない。マニュアルを読むことは当然として、音響分析に関する教科書の数冊は読んでから分析にかかる心構えが肝要である[1]。最近ではこの種の教科書も数多く出版されるようになってきた。一冊だけ推薦するならば Peter Ladefoged の Elements of Acoustic Phonetics の第二版(Chicago Univ. Press, 1996)だろうか[2]。

2 生理的データ

筋電図、X線、磁気、MRI などによる調音運動の観測、ファイバースコープによる声門の観察などのデータである。生理的データは音声研究上もっとも基礎的なデータであるが、本稿の読者がこれを分析する機会はありませんと思われるので、生理的データの収集とはどのような作業であるのかを理解していただくことを眼目として解説をすすめる。

生理的データの収集は音響データに比べてはるかに困難である。また手法によっては医師免許その他の資格が必要とされることもあるので、筆者は東京大学医学部音声研などとの共同研究によってデータを収集している。筋電図における筋肉への電極挿入や鼻腔へのファイバースコープ挿入などには一定の痛みも（皮下注射程度であるが）避けがたい。被験者は字義どおりの意味でのヴォランティアであることが多い。

また生理的データの収集は人体の解剖学的構造によって強く制約される。ファイバースコープによって声帯振動を観察する場合を例にとろう。光ファイバーは外鼻孔から挿入され、鼻甲介をすりぬけて開口部に達し、そこから下って咽頭の下部、声門を直上から観察可能な位置に固定される。この状態で奥舌母音や軟口蓋子音を調音すると、後退した舌はファイバースコープの視野をさえぎるか、場合によってファイバースコープに接触してしまう。そのため、実験の素材とする語や発話からは奥舌母音や軟口蓋子音などを注意深く除外しておかねばならない。

同種の問題は筋電図データの収集においても生じる。ピッチの下降に関与する可能性があると思なされている外喉頭筋 SH（胸骨舌骨筋）を例にとろう。この筋肉はピッチの下降時にも活動が観察されるが、同時に顎の開大や軟口蓋子音の調音などにも活動するため、実験の言語素材に奥舌面の運動に関係する分節音が含まれていると、観察された SH の活動がピッチ制御に関するものか、調音運動に関するものか判別が不可能になる。従って、言語素材はカ行子音を含まず母音も理想的には狭母音だけで構成された発話に限定されてしまう。無意味語を利用する実験ならば容易に企画できるが、有意味語でしかもある程度の長さをもった発話を構成するのはかなり難しい。筆者は数年前に日本語においてフォーカスの有無が語アクセントの実現におよぼす影響を SH その他の筋電図を用いて検討するために、上記の制約を満たした上に、アクセントの有無で対立する文セットをつくらうとしたことがある。『大辞林』の語彙を計算機を利用して悉皆的に検索しても適切な素材を構成することができず、結局は人名を利用して何とか切り抜けざるをえなかった[3]。

X線による舌などの調音運動の観察は音声研究全般にとって最大級に重要なデータであるが、被爆の問題があるためにデータ収集は非常に困難である。かつて国語研が撮影したX線映画などは、現在の被爆量に関する規制のもとでは撮影が許されない性質のものである。東大音声研で開発されたX線マイクロビーム装置

はこの問題を解決するために考案された測定装置であり、髪の毛よりも細く絞られたX線ビームを計算機制御することによって、被爆量を劇的に軽減させながら、舌や顎の表面に接着された金属球の位置変化を精密に記録することができる[4]。しかし、それでも被爆量はゼロにはならないので実験の総時間数が制約される。筆者が1996年に米国ウィスコンシン大学に設置されたマイクロビーム装置でみずから被験者となった体験では、半日かけて300発話程度のデータを収録すると、年間許容時間をほぼ消費してしまった。この装置によるデータ収集のもうひとつの難点は、虫歯の治療などで歯に金属補填剤が詰まっていると、それが金属球と同様X線を吸収するために、計算機が金属球の移動を追跡できなくなってしまう点である。

X線による被爆の問題を回避するために磁気による調音運動の観測手法も世界各地で研究されてきており、一部は商品化もされている[4]。日本でもNTT基礎研究所をはじめ数箇所稼働している。この方法では被爆の可能性は生じないが、舌表面に接着するコイルの位置が正中断面からずれると測定結果に大きな誤差が生じることが知られている。実験中のコイル位置を正確にモニターする手法が研究されているが、現在はまだ実用化されていない。また、この装置では被験者は頭部に巨大なヘルメット状の装置をかぶる必要があるため、頭部の運動をとまらざる種類が発話（例えば後述するパラ言語情報を含む発話）の測定には必ずしもむいていない。

3 知覚データ

音声研究には Articulatory, Acoustic, Auditory の三つのAがあるなどと言われることがある。前二者については上で触れたが、最後のA、つまり聴覚音声学（耳による主観的研究の意味ではない）と呼ばれる領域に関しては、残念ながら筆者は本格的な研究をおこなったことがない。また、音声知覚に関する本格的な研究はもっぱら実験心理学者の手によっておこなわれており、言語学の知見を活かした研究は僅少であることも事実である。管見のおよぶ限りで判断すれば、音声知覚研究で注意が必要なのは刺激の作成と実験計画である。

刺激の作成には音声の構造に関する知識が必要とされることは当然である。これについては後述する。音声の構造に関する知識とならんで必要とされるのが人間の聴覚に関する心理学的知識であるが、筆者はこれについて解説する能力に欠けているので、省略させていただかざるをえない。

実験計画に関して何よりも大切なのは実験しようとする仮説に関する要因の把握であり、ここでは研究者の言語学の知識が問われることになる。その他、通常の統計学的な配慮とならんで、実験の意図を被験者に簡単にさとられないための配慮も是非必要である。意図がみえみえの実験では、被験者が実験に協力的な反応を示す可能性が否定できないし、反対に性格の悪い被験者がいて、わざとデタラメな回答を試みないとも限らない（私のことではない）。

実験計画のファクターとして非常に重要なものに語の親密度(familiarity)がある。実験に用いる単語が平均的な日本人にとってどの程度親しみのある語であるかについての指標である。日本語の言語心理学的研究の重大な欠陥として、親密度に関する信頼すべきデータが存在しないことが長らく指摘されてきていたが、最近、NTT基礎研究所で8万語におよぶ単語親密度データベースが完成し、近く市販される予定である。今後の音声知覚研究にとって必携の文献となるだろう（詳しくは同研究所の天野成昭氏 amano@av-hp.brl.ntt.co.jp に問い合わせのこと）。

以上は、音声の知覚過程そのものを対象とした研究、いわば音声知覚の心理言語学的な実験についてであったが、音声研究ではその他にも種々の知覚実験を実施することがある。例えば音響的な分析で抽出された

パラメータの心理的な実在性を探るために、問題となるパラメータを操作した合成音を作成して、知覚実験を実施することがある。例えば日本語の分節音の長短に関する音韻的な対立を探るためにまずは分節音の持続時間を測定し、次に長短の閾値を知るために分節音の持続時間を加工した刺激音声を作成して、それらに対する長短の音韻的判断をもとめるといった研究である。この種の研究は音響的な分析結果の妥当性を保証するために是非必要であるのだが、そこまで踏み込んだ研究は、少なくとも文学部に属する研究者の間では、あまり実施されていない。その理由のひとつは、対象とする現象によっては一定の品質に達した合成音を作成すること自体が困難であることによると思われる。

概略的に述べれば、ピッチや音声振幅の加工はもっとも用意である。現在もっとも広く普及している音声分析手法である線型予測分析(LPC)によって抽出されるピッチや振幅のパラメータを手作業で加工してLPC合成をおこなえば、ほぼ満足のゆく品質を実現できることが多い。LPC合成にともなう音質の低下が問題となる場合は、PICOLAやSTRAIGHTのような技術も利用できる[]。

次に、持続時間の変更は対象とする分節音が定常的な性格をもっていれば容易である。無声閉鎖音の閉鎖区間、母音の定常部、無声摩擦音の(統計的に)準定常的なノイズなどである。ただし、母音のような有声音を波形の操作(cut/paste)によって延長・短縮する場合には注意が必要である。音声波形の周期をくずしてしまうと、それがスペクトルにひずみを生じさせるからである。波形の振幅がちょうどゼロになる点を選んで1周期分を選択し、それを単位として操作をおこなえば、この問題は回避できる。

最近、NTTインテリジェントテクノロジーから市販された「高品質音声合成メッセージ作成ツール」は、合成音のピッチ、振幅、持続時間長などをGUIを介して自由に操作できる面白いツールである。うまく利用すれば音声研究にも利用可能と思われる。ただし、テキスト音声合成(Text-To-Speech)ツールであるので、研究者の側で用意した音声を分析して加工するような用途には使えない。またフォルマントなどのスペクトル特徴の加工もおこなえない(URL: <http://www.ntt-it.co.jp>)。

一方、音声学で「わたり」と呼ばれる区間、つまりひとつの分節音から次の分節音にむかって時間とともに音質が変化してゆく部分を含む音声の加工には注意が必要である。単純に時間軸を引き伸ばしただけだと、わたりの時間特性が変化してしまうからである。わたりの時間特性が音声の知覚において重要な役割を果たしていることはよく知られているとおりでである。

フォルマントなどのスペクトル情報の加工はずっと難しくなる。スペクトル全体を周波数軸にそって伸縮させるのは比較的容易であるが、特定のフォルマント周波数だけを独立に加工しようとする途端に難しさがます。LPC分析で得られるフィルタ係数を加工することは可能だが、それによって目的を達するためにはデジタルフィルタに関する深い理解が必要であり、そういうことを実際にやっている研究者も私の周辺にはいない。

フォルマント情報の精密な制御が必要な場合、フォルマント合成(Formant synthesis)によって刺激音全体を作成することが考えられる。この場合、刺激音の全体に対して、5ないし10ms毎に合成用パラメータ(通常30個程度)をすべて指定する必要がある。定常母音の合成ならば比較的容易であるが、種々の子音や母音連続を含む音声を合成しようとする音響的な音声生成理論に関する理解にくわえて種々の名人芸も必要となり、挫折してしまう人が多い。高精度の音響分析が可能であれば、それによって得られた分析パラメータを初期値として合成をおこなうのが賢明だが、その場合でも、高精度の分析が可能な言語素材をあらかじめ慎重に選択しておく必要がある。例えば鼻音を含む音声は避けるべきである。LPC分析では原理的に鼻音の音響特徴を正しく抽出することが不可能だからである。LPCに代わる分析手法も種々研究されているが、現在のところ実用に供されていないから、言語素材の方で妥協するしかないのが現状である。また開発途上の手法を利用できる立場にある研究者にとっては本稿自体が無意味であるにちがいない。

II 研究事例

さて、以下ではふたつの研究事例を報告しよう。いずれも筆者が中心となっておこなった研究であり、現在も進行中の研究である。自分の研究であるから、事例として紹介するのは面映ゆいが、結局、筆者がもっともよく理解している研究を消化しすることにした。研究現場からのレポートとして読んでいただきたい。

1 パラ言語情報の伝達メカニズム

パラ言語情報とは言語情報に対する概念であり、話者の発話意図や心的態度などをさす。話者が意図的に表出している情報である点、および基本的には有限のカテゴリをなしているものと想定される点では言語情報（音韻情報）と共通しているが、単一カテゴリの内部で強さの連続的な変化が可能である点、および複数のカテゴリを同時に表出可能である点において、言語情報と異なっている。簡単に言えば音声を文字化すると欠落してしまう情報である。例えば「そうですか」というテキストには/so'H#desu#ka/という音韻情報が含まれているが、このテキストには「疑い」「感心」「落胆」「無関心」などさまざまなパラ言語的意味を込めて発話することが可能である。日常の音声コミュニケーションでは、テキストよりもむしろパラ言語情報を伝達することが重要であることも多いのだが、パラ言語情報が音声によってどのように伝達されているか、そのメカニズムはほとんど解明されてきていない。さて、このパラ言語情報をどう料理するか？

研究の開始にあたって、最初におこなったのは、日本語の教科書としてめずらしくパラ言語情報にも触れている Mizutani & Mizutani, *Aural Comprehension Practice in Japanese* (The Japan Times, 1979) の付属テープにおさめられた音声の音響分析であった[5]。これはあくまで、あたりをつけるための試行的分析であったが、当初耳で聞いて予想していたピッチおよび持続時間の変化以外に母音音質にも顕著な相違が認められるらしいことが判明したのは幸運だった。この段階で（さいわい所内の予算もとれたので）本格的な音響分析を想定したデータを収集することにした。

- 自発発話 (spontaneous speech) か 実験発話 (lab-speech) か
 - ⇒ 研究開始にあたってはデータの統制を重視して実験発話を用いる（ただし、最終的には自発発話の分析に立ち戻ることを目標とする）。
 - 下手な話者では元も子もなくなるので安定した発話が可能なお話者を探す。採用した話者は3名。いずれも日本語教師で音声教育に造詣が深く実用的な音声能力も高い人。うち1名は上記のテープを吹き込んだ話者その人である。
- データ収録上の注意
 - どのような特徴が関与的であるか予想がつかないので、とりあえずできるだけ高精度かつ多様な分析が可能なおデータを収録する。防音室でヘッドセットマイクを利用してDATに音声を収録。同時に声帯振動に関する情報であるEGG信号をDATの他チャンネルに記録。さらに話者の表情をビデオに記録。
- パラ言語情報の選択
 - パラ言語情報は「中立（棒読み）」「強調（声をはりあげた棒読み）」「感心」「疑い」「無関心」「落胆」を指定。Mizutani & Mizutani (1979) を参考に「中立」「強調」を追加。
 - （これとは別に「非難」「勧誘」「からかい」などについてもデータをとったが、今は省略）
- テキストの選択
 - 上記のパラ言語的意味を表出可能なテキストとして末尾が「Xデスカ」で終わるものに限定。
 - Xには、語の長さ・冒頭音節の音節構造・アクセントの有無と位置を考慮した要素を挿入。

ピッチ抽出が容易になるように、できるだけ有声音が連続する語を選択。

フォルマント情報の分析のために同一母音/a/が繰り返される語を選択。

● 発話数

分散分析などの統計的解析をおこなうために、ひとつの発話を最低 10 回繰り返す。

発話の順序はランダムイズ (ただし、ランダムには読めない話者もいた)。

● 分析

音声信号は DAT から直接計算機に入力。Sun-Solaris で稼動する esps+ xwaves という音声分析パッケージを利用して以下の作業をおこなった。

1. スペクトログラムの作成 ⇒ ラベリングのため。
2. 分節音のラベリング ⇒ 持続時間の分析のため (可能な限り音素単位に切る)
3. ピッチ抽出 (自己相関法により全自動)
4. ピッチの特徴点 (最高点, 最低点など) の測定および位置ラベリング
5. 特定の母音についてフォルマント周波数の測定 (DFT スペクトルを視察しながら LPC を適用。手作業)

以上の作業はアルバイトの大学院生が中心となって実施した。作業中に生じた疑義はつねにその場で筆者と相談して解決することとし、その記録を lab-note に随時記録していった。

音響分析の結果は試行的分析の結果が基本的に正しいことを示していた。また、新たにいくつかの問題も発見された。話者による個人差の存在、繰り返しにともなう同一話者内でのゆれの存在、発声様式 (phonation type) の変化などである [6-9]。個人差や個人内のゆれが問題になるのは、観察されたゆれがパラ言語情報を知覚するに際して許容範囲のうちにおさまっているかどうかという点である。一方、発声様式はいわゆる声質 (voice quality) の重要な因子であるが、その分析手法は現在でも十分に確立されているとはいいがたい。これについては新たな分析手法そのものの開発が要請されることになる。これらふたつの問題について、現在以下のような検討を進めている最中である。

まずゆれの問題については、以下の知覚実験を実施した。三つの文を選んで、そのすべての発話 (3 文 × パラ言語情報 6 種 × 話者数 2 ~ 3 × 繰り返し数 9 ~ 11 で 436 発話) を素材とし、被験者 20 名によるパラ言語情報タイプの同定 (identification) 実験を実施した。同定データは或る話者の「強調」を例外として、高い同定率 (全体の平均で 80% 以上) を示していた。同定データだけでは、上記の疑問が十分解決されたことにはならないので、さらに、この同定データから一定の手順によって 436 個の発話間の距離行列を計算し、それを多次元尺度構成法 (MDS) という手法によって解析し、刺激が分布する心理空間を構成した。その結果、6 種類のパラ言語情報は話者の別を越えて、明瞭に分離して分布していることが判明した。

さらに続いて、音響分析によって抽出された音響パラメータ群と心理空間との関係を分析した。心理空間 (3 次元) における 1 ~ 3 軸の座標値を従属変数、音響パラメータ群を独立変数とした重回帰分析をおこなうと、各軸とも高い精度での予測が可能であることが判明した [10]。これによって、抽出された音響パラメータに心理的実在性が存在することを、間接的に示すことができたわけである。今後は合成音声を利用して、各パラメータを独立に操作した刺激音声を作成して知覚実験を実施することによって、心理的実在性をより直接的に証明してゆきたいと考えており、実験を準備中である。

次に発声様式の問題については、科学研究費の補助をうけて、東大医学部および宇都宮大学工学部との共同研究をおこなっている。研究の主眼は、声帯振動 (音響的には喉頭音源) の特性を生理と音響の両面から

解明することにある。東大が開発した超高速ビデオ撮影装置(4500 frame/sec) を無響室内に設置して、定在波が存在しない状態での音声信号と声帯振動画像を同時収録し、両者を関連づけつつ研究をすすめている。第一報は今夏に発表の予定であるが[11]、「疑い」の冒頭では pressed voice が、「落胆」では breathy voice が観察されるなど、声帯の振動様式にはパラ言語情報と関連した顕著な相違が生じていることが生理・音響の両面から確認されつつある。

以上、この研究の過程をデータという観点からふりかえると、以下のような特徴ないし教訓を指摘できるのではないかと思う。

- 試行的研究の重要性（特に先行研究がない現象を研究する場合）
- 話者選択と発話条件選択の重要性
- 統計的検定を前提としたデータ収集
- 分析結果の知覚実験による検討（心理的実在性）
- 新しい分析手法の利用・開発，そのためのデータ収集

2 熊本方言イントネーションの研究

第二の事例として熊本方言のイントネーションの語用論的側面に関する研究をとりあげる。この研究は科研費重点領域研究『日本語音声』（代表者：杉藤美代子）の一部として偶然に開始したものであったが、その後、他の科研費の補助を得ながら、現在まで継続してきており、以下のような経緯で発展してきた研究である。

- | | |
|-------------------------------------|-----------|
| 1) 熊本方言イントネーションの音韻論的構造に対する仮説構築 [12] | 1990-93 |
| 2) 推定された構造の知覚実験による検討 [13, 14] | 1992-94 |
| 3) 音韻構造上の自由度に関する理論的検討 | (1994-96) |
| 4) 発話の丁寧さとの関係に関する知覚実験 [15] | 1996-97 |
| 5) 社会的変数を重視した知覚実験 [16] | 1998- |

1)の段階は、対象とする言語現象の構造自体が不明な段階である。また、そもそもこの段階では無アクセント方言である熊本方言にイントネーションの構造と呼べるものが存在しているかどうか不明であった。実際、この研究の出発点となった「日本語音声」の研究会では懐疑的な意見を表明される研究者が多かったように記憶している。しかし、これらの懐疑的な見解の根底には、イントネーションと無アクセント方言についての誤った見解が横たわっているように思われたので、調査研究を担当することに決まった当日、上記の研究会に出席していた熊本出身の吉岡泰夫さんを喫茶店に連れ込んで、疑問詞疑問文と単純疑問文のイントネーションなどについて、即席の調査を試みた。その結果は、明らかなイントネーション構造の存在をうかがわせるものであったので、日本語のイントネーションに関する文献を広く渉猟して、方言の差を越えて観察されそうな言語現象を列挙した調査票を作成することから研究を開始した。幸い 1980 年代後半には国内外でイントネーションに関する重要な研究が相次いで発表されていたので、大いに助かった。特に Pierrehumbert と Beckman による日本語イントネーションに関する実験音韻論的研究は重要であった。

熊本で数回にわたる臨地調査をおこない、収録した音声資料からピッチ抽出をおこなった結果、熊本方言のイントネーションの音韻構造についていくつかの仮説をたてることができた。熊本方言イントネーションの構造は、抽象的なレベルに関する限り、東京方言のそれとよく類似したものであること、ただし、熊本方言にはアクセント句のピッチの頂点をなすHトーンが当該アクセント句中のどの音節とむすびについてもよ

いという、東京方言には存在しない重要な特徴(wandering H と筆者は呼んでいる)があり、そのために表面上きわめて多様なイントネーションが実現されるという仮説である。また、ひとつの音節に結合される音韻論的トーンの数には上限があること、upstep という特殊な音声実現規則が存在すること、なども予想された(もうひとつ、熊本方言のイントネーションは、藤崎モデルと呼ばれるモデルでは生成できないという事実も重要な発見であるが[13], ここでは触れないことにする)。

この段階で、音韻論的なモデルから実際のイントネーションを合成する研究をおこなうことにした。上記2)の段階である。このような方向を選択した理由はいくつかあった。第一の理由は上記の wandering H である。表層音韻表示においてリンク先が決まっていなようなトーンは、管見の限りどの言語についても報告されていないし、また理論家のなかには、その種のトーンは表層にいたるまえに削除されるという制約を設ける人もいるから、そのようなトーンを心理的実在物として想定することの妥当性を実験的に示す必要があると考えられたのである。

第二の理由はいわば哲学的なものである。従来の音韻研究は当然のこと、いわゆる実験音韻論的研究でも現象の分析ばかりが中心となっていて現象の合成を試みる研究が僅少であることに筆者は不満を感じている。デカルトの「方法序説」の時代から自然科学の研究は分析だけでは終わらずに合成にまで進むことが欠かせないことは科学研究上の常識となっていると思うのだが、言語研究ではこれが必ずしも常識とはなっていないことに対する不満である。

そういうわけで、上記のモデルから生成される文法的なイントネーションと、何らかの形で非文法的なイントネーションとを多数合成し、それを熊本出身の被験者に聞かせて、イントネーションの自然さを評定してもらい知覚実験をおこなった。実験の結果は、文法的なイントネーションと非文法的なものとの間には自然性に明瞭な差が存在しており、特にアクセント句のフレーズが自然性に大きく影響することが示されていた。Wandering H に関しては、アクセント句のピーク位置が大きく移動しても音声の自然性はほとんど低下しないことが確認された。

このようにして、熊本方言のイントネーションの構造に関する仮説をたて、その妥当性を一応示すこともできた。しかし、wandering H に関しては疑問が残っていた。物理的には大幅に形状のことなる音声も音韻的には同一のカテゴリに属すると判断される現象は、異音現象としてどの教科書にも記載されている。熊本の wandering H も異音とみてよいのであるが、問題はそれがどのような異音なのかという点である。この問題に関する検討には長い時間を要した(上記の3))。初期の研究発表では、フォーカスとの関係を考えてみたりもしたが、どうもうまくまとまらない。分節音に関して、米国の Labov らは、従来の構造言語学的な音韻記述において自由異音とされてきた音声の使用実態が、社会的諸条件と明確な相関を示すことを明かにしてきている。これと類似した状況が wandering H についても存在するのではないかという可能性も考えられた。

実は、この問題は現在でもまだ解明されないままなのだが、1995 にひとつの仮説を思いついた。ピーク的位置によって発話の「丁寧さ」が変化しているのではないかという仮説である。これが上記4)の段階の研究になる。この研究は熊本方言自体の研究というよりも、より広い射程の研究の一部として位置づけられるものである。研究の趣旨は、発話の丁寧さに韻律が関与しているという直感的な事実を定量的な手法で検討することにあり、その素材のひとつとして熊本方言を利用しようというものである。

この実験では、いくつかの工夫をこらした。まず第一に韻律を操作した音声を被験者に聞かせて、丁寧さを何らかの方法で評定させた結果、丁寧さに変動が認められたとしても、その変動の大きさを何らかの規準に照らして評価できなければ説得力がないと思われたので、韻律と同時に語彙的要因も操作することにした。

この点で熊本方言は実に都合のよい言語である。もっぱら丁寧さの水準によって使い分けられる疑問の終助詞が3種類存在するので、それらを入れ替えることによって、語彙的な丁寧さを操作し、その変動と韻律による変動とを比較することによって、韻律が丁寧さにおよぼす影響を定量的に評価できるようにした。

第二に韻律的要因としてイントネーションをとりあげるのだが、発話末の上昇/下降の別と並んで、既に作成していた熊本イントネーションのモデルにしたがって、アクセント句のピーク位置も組織的に操作することにした。これによって、丁寧さの知覚におよぼす韻律の影響という実験本来の目的と同時に、懸案となっていた wandering H の問題についての解明もはかれるのではないかと期待したわけである。

最後に、発話の「丁寧さ」という連続的心理量に対して被験者に一貫した判断を要求することは、かなり困難であると思われたので、データの取得は一対比較法という形式によることにした。一対比較法では、ふたつの刺激を対として呈示し、問題となる属性に関してどちらの刺激のほうが勝っているかの判断だけを被験者に要求する。系列比較法のように属性の強さに関する評価を要求する方法に比べれば、被験者は比較的容易に判断をくだすことが可能であり、また反応の一貫性も高いことが期待できる。ただし、一対比較法には刺激数の増加について判断すべき対の数が急激に増加するという欠点もある。熊本の知覚実験の場合、イントネーション形状6種×終助詞3種で18種類の刺激を用いたので、刺激対の数は ${}_{18}C_2=153$ となり、刺激呈示の順序効果も考慮にいれてA・Bの対を念のためにB・Aの順序でも呈示したので、結局306対の実験となった。これを実施するには30~50分を要する。被験者の聴覚的な集中力が持続するギリギリの限界であろう。

実際に実施してみると、この実験の結果はなかなか興味深いものであった。まず、発話の丁寧さは終助詞によって組織的に変動し、そして、平均値および級内分散で評価するとそれとほぼ同程度、イントネーションによっても変動することが示されていた。また2元配置分散分析の結果は、語彙的要因とイントネーションとの効果は統計的に独立していることを示していた。

上でイントネーションと呼んだのは、ピーク位置と発話末の効果を込みにした6種類のイントネーションの効果である。それでは、ピーク位置単独ではどの程度の効果が存在するのだろうか。イントネーションを発話末の上昇/下降とピーク位置の特徴に分解して3元配置分散分析にかけてみた。両者の効果は統計的に独立しており、発話末の上昇/下降には単独でも高い有意性が認められたが、残念ながらピーク位置単独の効果は認められなかった(実は被験者グループ(後述)によってはピーク位置の効果が認められることもあったのだが、大きな差ではなかった)。丁寧さの伝達における韻律の関与とその定量的評価という実験の大目的は達成されたが、wandering H に関する疑問の解明は、さらに今後の課題として残されることになったわけである。

ところで、この実験では大学生と社会人(平均47歳)のふたつの年齢層から被験者を選んでいった。実験結果を子細に観察すると、ふたつのグループ間には反応に若干の相違が観察される。社会人グループの方が終助詞の効果が大きいのである。換言すれば、若い被験者の方が、よりイントネーションに依存して丁寧さを知覚しているわけである。これは社会言語学的に興味深い現象であると思われたので、より広い年齢層のデータを収集することにした(上記5)の段階)。具体的には、より年齢の高い社会人と中学生のデータを収集した。実験自体は前回の実験と同一である。

新しいデータは現在分析をすすめている最中であるが、現在までに分析が終了したデータは、上に指摘した傾向を一層あきらかに示している。平均年齢56歳の社会人グループは従前の社会人グループよりも終助詞に依存する程度がおおきく、反対に中学生は大学生以上にイントネーションに依存した判断をおこなっている。この問題の背景には、熊本方言の疑問終助詞のひとつ「ナ」が近年急激に失われつつあることと関

係しているようであるのだが、この問題は音声よりも社会言語学の範疇に属する問題であるので、今は省略することにしたい。

さて、熊本方言イントネーションに関する研究からは以下のような教訓をひきだすことができる。

- 海外もふくめて最新の研究動向を把握しておく必要性
- 言語学的な仮説構築の重要性
- 分析にとどまらず合成を試みる必要性
- 統計的検定を前提としたデータ収集
- 隣接分野（例えば心理測定法）に関する知識の重要性

ここで特に指摘しておきたいのは実験的研究における仮説構築の重要性である。筆者は最近、内外で音声関係の雑誌の査読にたずさわる機会が多いのだが、「実験」をタイトルに銘打った研究で、事実上、「観察」や「観測」の水準にとどまった研究が多いことを痛感させられている。観察や観測から「実験」が区別されるのは、実験が仮説の検証のためにおこなわれる作業であり点においてであり、仮説をもたない実験研究はありえない。そして、仮説は理論によって生み出されるのであるから、実験的研究を実施するためには、何らかの理論が存在していなければならない。理論的基盤が曖昧なままにおこなわれた実験は、それだけで価値が半減すると言ってよい。私は理論が嫌いだから実験をやっています、という研究者は本質的に矛盾をはらんだ言明をおこなっているのである。

観察や観測が科学研究の重要な1ステップであることは事実だが（パラ言語情報に関する筆者の研究の現状はこの水準にある）、それと「実験」の間に存在する質的な相違—つまり理論とそれによる仮説構築の有無—を正しく認識しておかないと、研究の正しい発展が阻害されるのではないかと危惧される。仮説をもたない科学研究は科学ではないと断言することはできないし、仮説は構築できてもそれを実験によっては直接検証することのできない科学の領域もある（進化論や宇宙論、言語研究ならば失語症や言語獲得の研究）。しかし、仮説を構築することが可能であるのにそれを実行しないのは研究者の怠慢というべきであろう。

III 研究事例のまとめ

観察＝現象の質的な把握

観測＝量的な測定にもとづく現象の把握

実験＝観察ないし観測に基づいて構築された仮説の客観的手法による検証

どの段階の研究をおこなうかによってデータをとる前提が異なってくる。観察で重要なのは包括性である。観測では観測対象の選択と測定手段の適否が問題になってくる。実験の段階にすすむためには、仮説の構築が不可欠であり、そのためには理論が必要である。最後に実験の企画と分析には、統計学・心理学・音響学・生理学など隣接領域の知識がしばしば必要になる。

研究成功の確率＝問題を発見する確率 * 問題を正しく分析する確率

問題発見能力 (serendipity)	知識＋理論＋幸運
分析手法	領域固有の手法の利用 隣接領域の手法の援用 新しい手法の考案

参考文献

- [1] 前川「音声研究者養成の課題」音声研究 1-1, 1997.
- [2] 前川「書評：Peter Ladefoged 著 Elements of Acoustic Phonetics」音声研究 1-3, 1997.
- [3] Maekawa et al. Electromyographic study of focus and accent in Japanese. *Journal of Acoustical Society of Japan (E)*, 16-5, 1995.
- [4] 特集「音声研究の新しい手法」音声研究, 2-2, 1998.
- [5] 前川「韻律によるパラ言語情報の表出に関する準備的考察」音声言語情報処 (情報処理学会資料) 13-6, 1996.
- [6] 前川「音声によるパラ言語情報の伝達：言語学の立場から」音響学会講演論文集 (秋季), 1997.
- [7] Maekawa. Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. *Proc. ICSLP98* (CD-ROM), Paper #0997, 1998.
- [8] 前川「音声によるパラ言語情報の伝達」国語研 50 周年記念研究発表会資料集, 1998.
- [9] 前川「韻律とコミュニケーション」日本音響学会誌, 55-2, 1999.
- [10] 前川・北川「パラ言語情報の知覚：同定データの多次元尺度法による分析」音響学会講演論文集 (春季), 1999.
- [11] Kasuya et al. Joint estimation of voice source and vocal tract parameters as applied to the study of voice source dynamics. *Proc. ICPH99* (In press).
- [12] 前川「無アクセント方言のイントネーション (試論)」音声言語, 4, 1990.
- [13] 前川「アクセントとイントネーション—アクセントのない地域—」杉藤 監修『諸方言のアクセントとイントネーション』三省堂, 1997.
- [14] Maekawa. Intonational structure of Kumamoto Japanese: A perceptual validation. *Proc. ICSP94*, 1994.
- [15] 前川・吉岡「発話の丁寧さに対する語彙的要因と韻律的要因の寄与」国語学, 190, 1997.
- [16] Maekawa. Contributions of lexical and prosodic factors to the perception of politeness. *Proc. ICPH99* (In press).