

音声研究のための録音入門 —再現性を保証するために—

菅井 康祐

近畿大学

キーワード: 音声, 録音, 機材, 設定, 手順

1. はじめに

外国語教育の分野でも学習者の発話や、聞き取りを調査したいと考える研究者・教育者の方は少なくないであろう。しかし、音声の研究には様々な機器や知識が必要であり、敷居が高いと感じている方も多いと思われる。確かに、音声の録音には手間がかかるのは事実であり、信頼の置ける録音をするにはそれなりの機器および手順が必要となる。しかし、以前に比べると音声の録音・分析に必要な機材はかなり安価になり、その操作についても以前のような職人的な技術がなくても十分使いこなせるようになってきている。

その一方で、音声の研究、とくに外国語教育においては、音声に現れる音響的な特徴が言語としてどれほど意味のあるものかという理論がまだ確立していないのも事実である。言い換えれば、音声はその受け取り手がどのように知覚・認識するかという問題とは切っても切り離せない関係にあるので、認知的な理論の枠組みが変われば音響データの解釈も変わりうる。このような点を踏まえて、そのよって立つ枠組みが変わっても再分析に値するような信頼できるデータを残しておくことが重要である。そこで本稿では、研究対象として音声を録音する際に気をつけておきたい点を、最小限に絞り具体的に紹介したい。

2. 録音機器

音声の録音において重要なことはいかに S/N 比 (signal-to-noise ratio) を上げるかということに尽きる。つまり、可能なかぎり調査対象となる音声情報のみを記録し、その他の音 (ノイズ) を情報に乗せないかということが重要である。そのためには、適切な機材を選ぶことがまず重要になる。

2.1 集音

音声は、空気を伝わる振動・波であるという性質を考え、上記の S/N 比を上げるためには、とにかく音声以外の音・振動をマイクに伝えないことが重要になる。また、インフォーマントが直接接するのはこのマイク周りの機器になるので、インフォーマントの負担をできるだけ減らすということにも配慮しなくてはならない。これらの点から集音機器の選択、及び設置についての基本的な考え方を紹介する。

マイク

マイクには大きく分けて(エレクトリック)コンデンサ・マイクロフォンとダイナミック・マイクロフォンの2種類がある。詳細なマイクの構造等についての説明については、平坂(2009)などを参照いただくとして、ここでは両者の大まかな特徴を表1にまとめる。ただしこれはあくまで一般的な比較であり、当然個々の機器の性能によってはこの比較に当てはまらない場合もある。結論から述べると、コンデンサ・マイクは小さな音量に対しても感度が高いため、固定した状態での発話・楽器演奏や、屋外での環境音録音などに向いている。それに対し、ダイナミック・マイクは頑丈で、大きな入力にも耐えられるので、コンサートなどのボーカルマイクに向いている。

表1 コンデンサ・マイクロフォンとダイナミック・マイクロフォンの比較

	コンデンサ・マイク	ダイナミック・マイク
大きさ	小さい	大きい
丈夫さ	弱い	強い
感度	高い	低い
価格	高価	安価

このような特性を踏まえると、発話音声の録音には、多少高価ではあるが、コンデンサ・マイクを使用し、さらに、不必要な音声をできるだけ排除するために、単一指向性¹のものを選びたい。また、マイクによっては、低音に対する感度が高い物や高音に対する感度が高い物など、周波数特性が異なる。分析対象として音を記録する際にはできるだけ原音を忠実に記録する必要があるので、マイクカタログなどを参照し、できるだけ周波数特性が平坦(フラット)な物を選ぶようにしたい。

マイク設置器具

マイク選びと共に重要なのが、マイクスタンドの選び方・設置である。実際にインフォーマントの録音中の精神的・肉体的負担を決定づけるのはマイクスタンドであると言っても良い。理想的なマイクスタンドとは、マイクに全く振動を伝えないほどしっかりとっていて、それでいてインフォーマントにはマイクの存在を全く感じさせないようにセッティングできるものである。いくら質の良いマイクを選んでも、スタンドからノイズの振動が伝わっては意味がないし、マイクの存在に意識が向きすぎ、発話に影響が出てしまうのも問題である。具体的なスタンド選びのポイントとしては、まず、床や机などから伝わってくる振動を押さえ込むだけの十分な重量があり、可動部(ボルト、ネジなどで留められている部分)が、しっかりと固定できること。次に、実際に録音する場所を考え、最適な位置に最適な角度にマイクを設置できることがあげられる。マイクスタンドには主に、卓上型とブーム型の2つのタイプがあるが(図1a, 図1b)、使用環境に応じて適切な物を選択したい。なお、録音に慣れていないインフォーマントにとっては、マイクの前に座って声を録音されるだけでもストレスがかかるので、万全の準備をした上で、実際に実験者がマイクの前に座り、姿勢に負担がかからないか、視覚的に圧迫感を感じないかなどを何度も確認したい。



図 1a 卓上スタンド



図 1b ブームスタンド

ウインドスクリーン(ポップブロッカー)

コンデンサ・マイクを用いて音声を録音すると、その感度の良さのために、様々なノイズの影響を受けてしまう。特に声を発するときマイクに当たる風圧(吹かれ)の影響は非常に大きい。これを避けるためにマイクと口の間に設置するのがウインド・スクリーン(ポップ・ブロッカー)である(図 1a のマイクの右側の器具)。これは、細かい網の目状になっており、無駄な空気の流れを遮ってくれる。図 2a・2b の音声波形を比べると、ウインドスクリーンを用いずに録音した音声の波形は全体の音声信号が小さいにもかかわらず、それぞれの波形の頭に吹かれによる大きな波が現われているのがわかる。なお、ウインドスクリーンを用いても吹かれの影響が出てしまう場合は、マイクの角度を変え、口からの風圧をマイクが直接受けないようにセッティングする²。

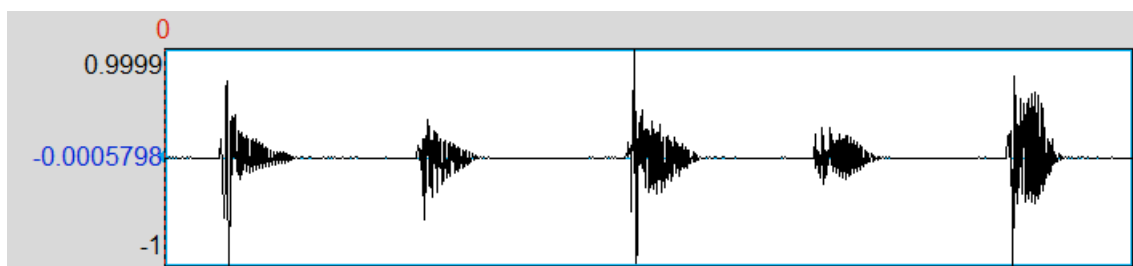


図 2a ウインドスクリーンを使わずに録音した日本語の「パ」「ピ」「プ」「ペ」「ポ」の音声波形

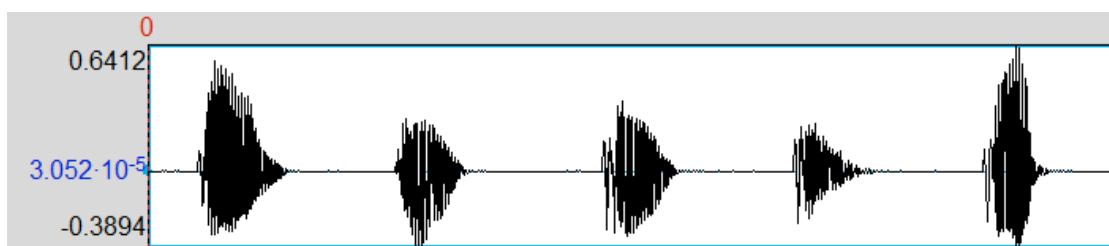


図 2b ウインドスクリーンを使って録音した日本語の「パ」「ピ」「プ」「ペ」「ポ」の音声波形

2.2 記録装置

マイクで集めた音声を記録する装置にも色々な物がある。以前は、テープレコーダなどのアナログ機器が中心であったが、DAT (Digital Audio Tape) や MD (Mini Disk) の登場により、デジタル録音を中心になった。現在では HDD などの記憶媒体の容量も格段に大きくなっているため、データはデジタルで、可能な限り高品質で残したい。デジタルで録音できる機器といっても様々であるが、音声を圧縮した形で記録する MD (すでに持っている人が少ないと思われるが) や、駆動部をもつ DAT などは、メカノイズを伴う可能性があるため避けた方が無難である。

ここで有力な選択肢になるのが、メモリ・レコーダやパソコンである。筆者としてはメモリ・レコーダの使用を勧めたい。まず、記録媒体がフラッシュメモリなどであるため駆動部を持たず、無駄なメカノイズがない。また、小型で携帯しやすいので場所を選ばないのも長所である。ただし、品質については価格に応じて大きな差があるので、最低でも非圧縮形式 (PCI, wav)³ のファイルでそのまま記録できるものを選びたい。

次の候補であるパソコンを記録装置として使うにはいくつか注意が必要である。まず、パソコンそれ自体が電磁波を発生しているため、市販の状態のままマイクをつないで録音するとほとんどの場合、ノイズが入ってしまう (図 3a)。それを避けるにはサウンドカードを質の高い物に載せ替えるか、USB 経由の外部サウンドポートなどを通してマイクを接続する必要がある。また、パソコンのファンのモータによる振動もノイズになるので、マイクとパソコンは十分な距離を置く必要がある。図 3a・3b を比較すると、サウンドカードの違いによって母音「ア」の前の無音区間のノイズの大きさの違いが見て取れる。

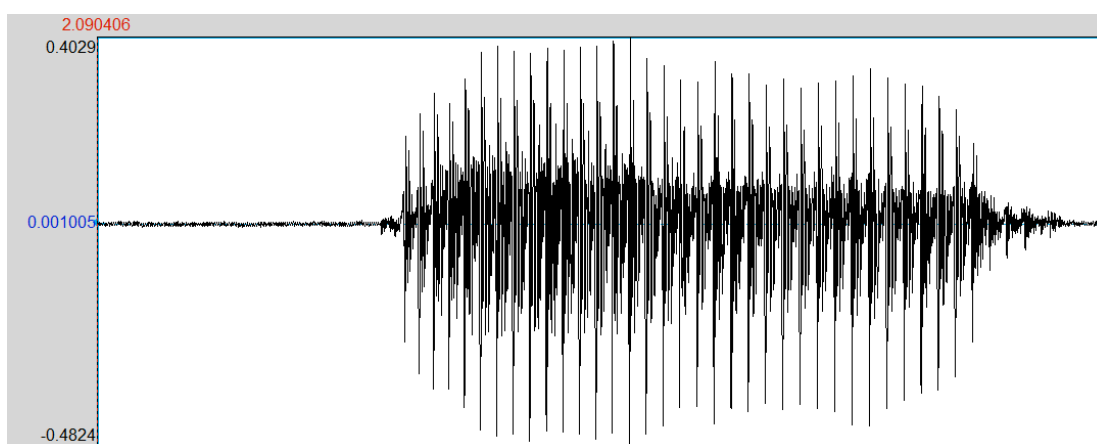


図 3a 富士通社製 D5230 パソコン純正サウンドカードを通して録音した「ア」の波形

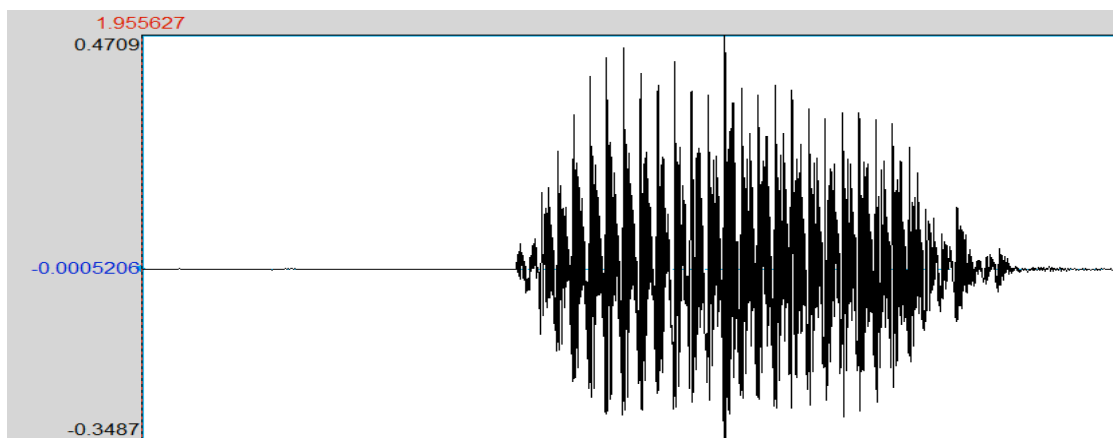


図 3b サウンドカードを ONKYO 社製サウンドカード SE200 に換えて録音した「ア」の波形

3. 設定

3.1 デジタル化設定

マイクで集められた音声はもとの音波がアナログで入力された状態であるので、前出のデジタルファイルに変換する必要がある(A/D 変換)。この過程で失われた情報は二度と復元することはできないので、可能な限り多くの情報量を残したい。ここで録音した音声をリスニングテストに用いる場合も、容量の問題で多少音質を落とさざるえない場合があるかもしれないが、元の音声はできるだけ高音質で保存することをおすすめしたい。

図 2a/b, 3a/b の音声波形のサンプルをみると、波は 2 次元で表現されるのがわかる。この X 軸の時間軸方向の情報量が周波数(Hz: 音の高さの情報)と大きく関わり、Y 軸方向の情報量が音圧(dB: 音の強さの情報)に関する。音のデジタル化の際に設定する値は、この X 軸の解像度を定めるサンプリング周波数と Y 軸の解像度を定める量子化(bit⁴)の決定、およびチャンネル数(モノラル/ステレオ)の選択である。このそれぞれの値に持続時間長を乗じた物がデータの大きさとなる。ちなみに CD の音質は以下のように表される。

1 秒あたりのデータ量

$$44,100\text{Hz} \times 16\text{bit} \times 2 = 1,411,200 \text{ bps (bit/second)}$$

(サンプリング周波数) (量子化) (チャンネル数:ステレオ)

74 分 CD のデータ量

$$1,411,200 \times 60 \times 74 = 6,265,728,000 \text{ b (bit)}$$

$$6,265,728,000 \text{ b} \div 8 = 783,216,000 \text{ B (byte}^5\text{)}$$

$$= \text{約 } 783 \text{ MB}$$

サンプリング周波数は、理論上設定値の半分の周波数の高さの音まで記録できるので、人間に聞こえる音の高さ(可聴域⁶⁾)から判断しても 44,1 kHz の設定で十分であろう。実際にはもう少し低い設定でも十分な品質ではあるが、前述のような理由から可能であれば高品質で保存したい。また、音圧についても CD 音質の 16bit で十分な解像度が得られる。ただ、チャンネル数については、音声をステレオ録音する必要は無いので、モノラル設定にすることでデータ量は半分になる。ただし、オーディオ CD として再生するには CD 規格に基づく必要があるので、CDR に焼き付ける際にはステレオに戻す必要がある。このような録音設定は、無料の音声録音・編集ソフトでも簡単にできる(図 4)。

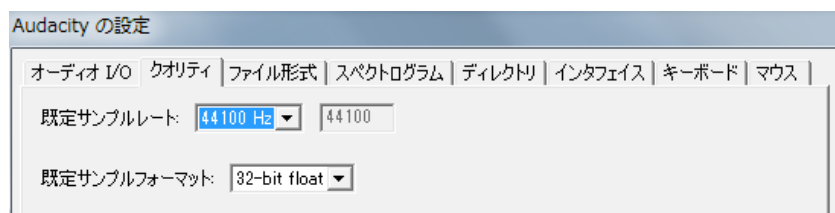


図 4 Audacity⁷ のサンプルレート(周波数), 量子化(サンプルフォーマット) 選択画面

3.2 録音レベル

録音レベルの設定が S/N 比をあげるために最も重要な設定である。データを記録できる範囲(クリップレベル)を超えない範囲でできるだけ大きな音源を録音できるように設定したい。録音機器、や録音ソフトで録音レベルをモニターしながら、最大の音圧の時に -3dB あたりになるように設定するのが一般的だが、機器のセッティングが終わったら、録音レベルをモニターしながら実際に何度かマイクに録音をして確認し、適切なレベルに調整することが望ましい。

3.3 録音環境

録音を行うには無響室、防音室、録音スタジオなどを利用するのが望ましいが、そのような環境が無い場合は、できるだけ音の入ってこない密室で静かな時間帯を選んで実際に録音してみると、波形上に突発的な雑音が入らないようであれば、録音は可能である。しかし、録音中はしっかりとインフォーマントの横に付き添い、ノイズが入った場合にはその録音セッションが終わったら必ず波形と音声で確認し、ノイズの混入が確認された場合は速やかにそのデータは破棄し、必要であれば録りなおす。

4. 録音時の注意点

ここまで技術的な面を中心に話を進めてきたが、実際録音の際にはすべてを決定するのはインフォーマントの声である。そのため、できるだけストレス無く心地よく発声できる環境を保つ必要

がある。そのためには、先に述べた機器のセッティングもさることながら、インフォーマントへの気配りが重要である。第一に、インフォーマントが自然に声を出しやすく、疲れにくい体勢をとれているか、座席・マイクの位置などをしっかりと確認する。つぎに、声は案外疲れが現われやすく、すぐに質が落ちるので、こまめに休憩・水分補給をとってもらおう。そのためには、実験の課題作成段階で、少なくとも5分に1度は休憩を取れるように考えることが必要である。また、録音のはじめは声の出方が不安定であったり、緊張があったりすることが多いので、課題のはじめには少しダミー課題を挿入した方がよい。また、課題の数があらかじめインフォーマントにわかっている場合は終わり間際で力が入ったり、気が抜けたりと声質が変化することもあるので、この場合には課題の終わりにもダミーを挿入しておき、あらかじめダミーを含めた数をインフォーマントに伝えておくのも一つの手段である。

また、リスニング用の音声を録音する際は、聞き手が聞きやすい声のナレータを選び、声質の影響を排除するために複数の話者の声を使用するべきである。

5. おわりに

本稿では、音声研究に用いる音声を実際に録音する際の機器の選び方・設定・手順などについて述べてきたが、録音になれていない読者の方には、「やはり音声研究は手間がかかる。」という印象を与えたかもしれない。しかし、しっかりと確信の持てるデータを残しておけば、何度でも異なる視点からデータを見直し、いつでもデータに語らせることができる。確かに、近年ではCALL 教室などで一斉に録音した音声を用いて統計的に処理することにより、一般傾向を推測するという手法もある。しかし、このような手法では、一人一人の録音状態にどのような要因が影響しているかの判断は難しく、どうしても状況依存性の高いデータになる。その結果、個々のデータの信頼度は低くなってしまう。母語の方言差・性格・学習経験など多くの個人差を含む音声データを扱うには、少人数のインフォーマントを対象に、個々からえられるデータの再現性を確認するという手法も重要であり、このようなデータから得られる物は大きいはずである。また、再現性の観点から考えれば、他の研究者の調査結果を再検証するだけでも有意義であり、自身の調査結果についてもいつでも他の研究者が追実験できるだけの情報(実験環境・使用機器・インフォーマントの情報)を開示し、お互いに研究成果を積み上げ、前進させていくことが重要である。そのための信頼のおける音声データの収集という点で本稿が少しでも役に立てれば幸いである。

注

1. 全方位の音を均一に集めるのではなく、マイクの向けられた狭い角度の音を集めるのに適した特性を持つ。
2. このほか、音楽録音などでは、他の楽器の音などがマイクに伝わるのを防ぐためにショックマウントという器具をマイクスタンドとマイクの間に着用することもあるが、実験環境での音声録音で

はそこまではする必要がないであろう。

3. MP3 ファイルなどは独自の方式で音声データを圧縮するので研究使用は避けたい。
4. bit はデータ量を表す単位で 2 の乗数である(例: 16 bit = 2^{16})
5. byte はデータ量を表す一般的な単位で 8 bit = 1 byte となる。また bit は小文字 (b), byte は大文字 (B) で表す。
6. 一般的には, 20–20,000Hz ぐらいだと考えられている。
7. web 上に公開されている無料の音声録音・編集ソフト(<http://audacity.sourceforge.net/>)。

参考文献

- 相澤昭八郎・高和元彦・半田健一 (1991). 『録音制作入門』 オーム社.
- 平坂文男 (2009). 『実験音声学のための音声分析』 関東学院大学出版会.
- 城生佰太郎 (2008). 『実験音声学入門』 サン・エデュケーショナル.
- ムーア, B. C. J. 大串健吾 (監訳)(1994). 『聴覚心理学概論』 誠信書房.
- ライアルズ, J. 今富摂子・荒井隆行・菅原勉 (監訳)(2003). 『音声知覚の基礎』 海文堂出版.