国立国語研究所学術情報リポジトリ

会話の収録・音声動画データの加工

メタデータ	言語: Japanese 出版者: 研究社 公開日: 2026-07-24 キーワード: 作成者: 臼田, 泰如 メールアドレス: 所属: 静岡理工科大学
URL	https://doi.org/10.15084/0002000722

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

第2章｜会話の収録・音声動画データの加工

本章では、会話コーパスにおける音声や動画の収録について述べる。音声や動画を適切に収録することは、コーパスを活用した研究の可能性や利便性を広げるだけでなく、収録後の転記作業などの効率向上にもつながる。そこで本章では、『日本語日常会話コーパス』の収録法に限定せず、収録機器を選定する際に必要となる情報や実際に収録する際の注意点、また収録後のデータの扱いなどについて広く扱うこととする。1節と2節では音声と動画の収録について、3節ではメタ情報の収集について、4節では収録後の音声・動画データの加工について概説する。

□1　音声の収録

1.1　録音機器の種類

防音室での録音といった実験環境下ではなく、**『日本語日常会話コーパス』**（**CEJC**）のように日常場面で交わされる会話を収録することを想定した場合、音声を収録する方法としては、ICレコーダーなど録音専用の小型の機器を利用する場合と、ビデオカメラに内蔵されているマイクを利用する場合とに大別される。

ICレコーダーなどの録音専用機器は、収録の状況や研究目的に合わせ、録音方式やマイクの指向性などを調整できることが多い。ただし、動画も同時にビデオカメラで収録する場合、別々に保存された音声と動画を同期させる必要がある。

ビデオカメラのみで動画と音声を同時に収録する場合、機器の操作が簡便で同期処理も不要なため、後処理の手間が少ないという利点がある。またビデオカメラを持って撮影者が移動するような場合にも適している。ただし、録音専用機器ではないため、音声設定を細かく調整できないことも多い。また話者から離れて設置することが多く、後述する超指向性マイクなどを用いても、話者の近くに録音機器を置いた場合と比べて音質が劣ることが多い。

1.2 録音機器の設定

音声を明瞭に収録することは、研究の幅を広げ、転記作業の効率も高める。しかし、必要以上に高精度で録音するとファイルサイズが大きくなり処理時にソフトウェアの動作が重くなることもある。そこで本節では、収録する会話の種類や環境に応じて適切な機器を選ぶために必要な知識についてまとめる。

■録音方式　デジタル録音方法には、**リニアPCM**などの非圧縮方式と、MP3などの圧縮方式がある。リニアPCMは原音からの損失が小さいが、ファイルサイズは大きい。一方、圧縮方式はデータが欠落するため音声分析には不向きだが、ファイルサイズを抑えることができる。ビデオカメラでは、録音形式がMP3などに固定され変更できないことが多い。一方ICレコーダーなどの録音専用機器では、リニアPCMを含む複数の録音方式から選択可能な機種も多い。

■録音レベル　録音レベルとは、録音における音量の感度を指す。主に声の大きさや**マイク感度**、話者との距離によって収音音声の大きさが決まるため、小さい声や遠距離の音声を録音する場合は入力レベルを相対的に高めに設定する。録音レベルは**デシベル**（dB）で表され、最大値を0dBとする負の値をとる。レベルを上げすぎると0dBを超えてクリップ（歪み）が生じる一方、低すぎると**ノイズ**が相対的に大きくなる。

■サンプリングレート　アナログ音声をデジタル化する際、**標本化**（**サンプリング**）を行う。1秒間の標本数（周波数）がサンプリングレートである。民生用の録音機器では44.1kHzや48kHzが一般的である。数値が大きいほど音が滑らかで高音域の表現に優れるが、ファイルサイズも大きくなる。人の可聴域を踏まえ、収録時は44.1kHzか48kHzに設定することが多い。

■量子化ビット数　サンプリングでは、音声は離散的な値に変換される。この変換を**量子化**と呼び、量子化ビット数はその値の細かさを示す。ビット数が大きいほど**ダイナミックレンジ**（音量の最大値と最小値の幅）が広くなる。民生用録音機器では16bitが一般的だが、24bitに対応した機器もある。日常会話のダイナミックレンジは約60dBで、16bit（96dB相当）でも十分対応可能である。ただし、例えば小さな声を録音する場合は、24bit設定によってより細かく音の強さを記録でき、小さな音の解像度が高まるなどの効果がある。

■マイクの指向性　マイクの指向性は大きく分けて**単一指向性**と**無指向性**（**全**

指向性）がある。単一指向性マイクは特定の方向からの音に対して感度よく収音するのに対し、無指向性マイクは全ての方向の音を収音する。単一指向性よりも指向性が鋭いものを**超指向性・鋭指向性**マイクと呼ぶ。**ガンマイク**や**ズームマイク**と呼ばれることもある。会話全体を録音する場合には無指向性を、特定の話者の音声をねらって録音する場合は単一指向性や超指向性を選択するといったように、対象や状況によって使い分ける必要がある。

ICレコーダーには通常、2つ（機種によっては3つ以上）の内蔵マイクが搭載されており、音の方向性や空間の広がりを再現することができる。機種によっては、マイクユニットの角度を調整して収音範囲をコントロールしたり、前方の音を強調するズーム録音機能を備えているものもある。

1.3 実際の録音

■収録環境　会話を録音する際には、収録環境に注意を払う必要がある。日常生活には、洗濯機や車の走行音、屋外スピーカーなど、録音の明瞭性を損なうさまざまな雑音がある。録音場所を選べる場合は、雑音の少ない場所や静かな時間帯を選ぶことが望ましい。空調の送風音やパソコンの冷却ファンのように、普段は気にならない音も録音時には思いのほか大きく記録されることがある。これらを完全に排除することは難しいが、空調の吹き出し口の真下を避ける、不要なパソコンの電源を切るといった工夫で雑音を軽減できる。録音場所の残響にも注意したい。例えば床がタイルやコンクリートなどの硬い素材でできている部屋、天井が高く床面積の広い空間では残響が強まりやすい。一方、カーペットやカーテンがある小さな部屋では音が吸収されやすい。

■録音機器の配置　録音機器の設置にも注意が必要である。機器やマイクは雑音の発生源からできるだけ離して配置する。空調の近くで録音せざるを得ない場合、風が直接当たらない位置に設置したり、**ウインドスクリーン**を付けたりすることで、風によるノイズを軽減できる。また録音機器を話者に近づけることで、周囲の雑音に対して音声の入力レベルが相対的に高くなり、明瞭に録音できる。更に、録音機器を机に直接置くと、カップを机に置いたり机を指で叩いたりした際の振動が大きなノイズとして記録されることがあるため、マイクスタンドを使用するか、机にクッション材を敷いて機器を設置するとよい。

■録音レベルの設定　会話を録音する際には、録音レベルを事前に適切に設定することが重要である。多くの録音機器には、入力音声の大きさを視覚的に確

認できるレベルメーターが備わっている。会話中は声の大きさに変化があり、特に盛り上がった場面では音量が大きくなる。録音レベルは、やや大きな声で話した時にメーターが－12dBから－6dBの範囲に収まるように設定すると、音割れを防ぐことができる。機種によっては**マイク感度**という名称で録音レベルを実質的に調整する場合がある。小さな声や話者が遠い場合は、感度を上げて対応する。

ICレコーダーには、自動で録音レベルを調整する機能を備えた機種も多い。発話の音量に大きな変動がある場合などは、この機能により音量のバランスを一定に保つことができる。ただし、音量差が平準化されるため、分析時に本来の音量差が把握しにくくなる点には注意が必要である。

■2人会話の音声を収録する場合　2人の話者による会話の音声を収録する場合、マイクの指向性を広く設定したICレコーダー1台を2人の中間に設置するだけでも録音は可能であり、機材の操作も非常に簡便である。この方法では、いずれの話者の音声も1つのファイルに混在して記録されるが、インタビューなど発話の重なりが少ない場面では、十分に明瞭な音声が得られる。

一部の機種では、内蔵マイクのユニット角度を調整できる設計となっており、左右の話者にそれぞれのマイクを向けるよう角度を設定することで、相互の音声の混入をできるだけ抑えて収録することができる。

各話者の音声をより明瞭に分離して録音したい場合や、話者が移動する場合などは、**ワイヤレスマイク**の活用が効果的である。例えば、筆者らの研究グループが扱っている機材を例にとると、図1に示すように、Rode社のWireless GOシリーズのような2つのワイヤレスマイク（送信機）と1つの受信機をセットで運用できるワイヤレスマイクシステムをICレコーダーの外部マイク端子に接続することで、両話者の音声をある程度独立して収録することができる。図2に、この構成で収録した音声波形の例を示す。この例では、2人の話者が「おはようございます」と交互に発話しているが、波形を見ると、完全な分離とは言えないものの当該話者の音声がより明瞭に録音できていることが分かる。



図1　ワイヤレスマイクの接続の例

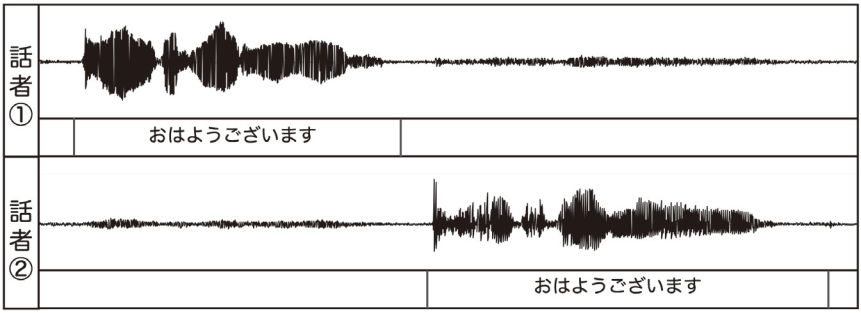

図2　ワイヤレスマイクで収録した音声波形の例

■3人以上の音声の会話を収録する場合　3人以上の話者が円卓を囲んで会話する場面を収録する場合を考えてみよう。2人会話と同様、ワイヤレスマイクを利用する場合、3チャンネル以上のマイク入力に対応した多チャンネル録音機器を使う方法が考えられる。図2で見たように、個々の話者の音声がより明瞭に収音されることが期待できる。しかし、多チャンネル録音機器は一般のICレコーダーと比べると大型で、マイクやケーブルなどと合わせると設置が煩雑で操作も複雑になるという課題がある。

これとは別に、各話者の前に個別のICレコーダーを配置して録音する方法もある。個々の音声ファイルの同期作業が必要となるが、機器の操作は簡便で実用的な方法である。第1章3節で言及したように、CEJCでは多様な場面における多様な話者との会話を一般の協力者に収録してもらうため、(1) 幅広い世代の一般の協力者が容易に操作できること、(2) 屋外での使用も想定し、携帯

性に優れた軽量な機材であること、(3) 話者の視界に入る位置に配置しても圧迫感を与えないことを機材選定の要件として掲げた。そのためCEJCでは、単一指向性に設定したICレコーダーを首から下げたフォルダーに収納し、口元近くに配置する方法を採用した（図3）。この方法で収録した2台のICレコーダーの音声波形の例を図4に示す。

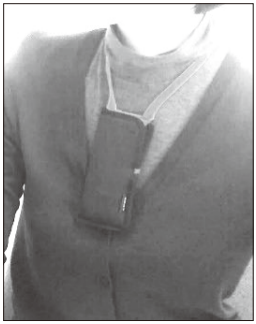

図3　CEJC収録時におけるICレコーダー装着の例

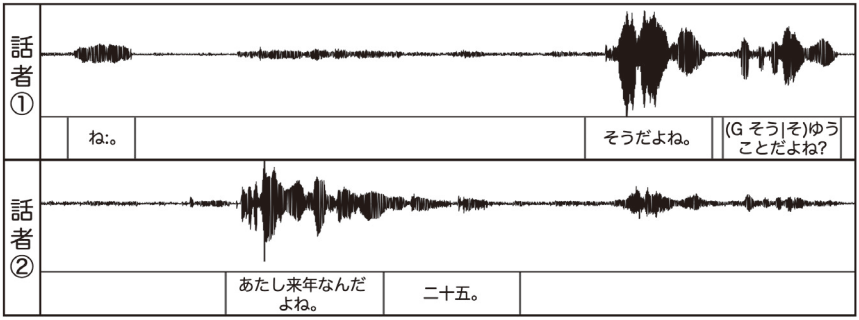

図4　2台のICレコーダーで収録した音声波形の例

□2　動画の収録

2.1　録画機器の種類

2025年現在、民生用ビデオカメラとしては大きく2通りの選択肢がある。一般に**デジタルビデオカメラ**と呼ばれるタイプと、いわゆる**ウェアラブルカメラ**と呼ばれるものである。前者は後者に比べて大きくて重量があるが、容易にAC電源で稼働させることができ、大容量のバッテリーも運用しやすい。また

一般にズーム機能があり、設置場所による画角や撮影範囲の制約の問題をある程度回避することができる。ただし本体重量や重心の問題から、大型の三脚に装着して収録することを前提とした構造をとっており、狭い場所では運用しにくい。一方、ウェアラブルカメラは小型で軽量であり、本体のみで防水機能を備えている場合もある。一方でバッテリーの持続時間は相対的に短く、大容量バッテリーのオプションは利用できない場合が多い。またAC電源での運用にも制約がある場合がある。加えて、ビデオカメラの音声を利用する場合、ウェアラブルカメラで収録できる音声には制約がある場合が多い（後述）。表1にこれらの比較をまとめた。なお、記録容量については、2025年現在ではどちらのタイプでも外部記録媒体を利用することが一般的になっており、両者の間に差は見られない。

表1　民生用ビデオカメラの比較

	デジタルビデオカメラ	ウェアラブルカメラ
サイズ	大	小
バッテリー	長（オプション利用可）	短
AC電源	標準的に利用可	制約あり
ズーム	標準装備	機能なし
音声	比較的高	比較的低
防水	なし	あり

2.2　録画機器の設定

■解像度　解像度とは、画像内に縦横どれだけ多くの**画素**（**pixel**）が並んでいるかを示す指標であり、高いほど精細で高画質な映像が得られる。2025年現在、多くのビデオカメラが4K撮影に対応している。4K（3840×2160画素）は、従来主流だったフルHD（1920×1080画素）の約4倍の画素数を持つ。解像度が高くなることで映像はより鮮明になるが、ファイルサイズの増加や編集・分析時の処理負荷の増大といった注意点もある。そのため、用途に応じて適切な解像度を選択する必要がある。

■フレームレート　フレームレートとは、動画を構成するフレーム（静止画像、いわゆる「コマ」）が1秒間に何枚表示されるかを示す指標で、fps（frames per second）という単位で表される。2025年現在、多くのデジタルビデオカメラで

は30fpsや60fpsでの撮影が可能である。フレームレートが高いほど動きのある動画がより滑らかに表示されるため、滑らかさを重視する撮影では高いフレームレートが好まれる傾向にある。一方、フレーム数が増えると1秒あたりのデータ量も増加し、ファイルサイズも大きくなる。そのため、被写体の動きや撮影の目的に応じて適切なフレームレートを選択することが重要である。

■視野角（FOV: Field of View） 視野角とは、カメラが一度に撮影できる範囲の広さを示す角度であり、広いほどより広範囲を一度に写すことができる。一方、視野角が狭い場合、撮影範囲は限定されるが、被写体をより詳細に捉えることができる。視野角は主にレンズの焦点距離に依存し、一般には35mmフィルム換算で「12–39mm」などのように表示される。視野角に直接影響するのは最小焦点距離であり、一般に最小値が小さいほど広い角度での撮影が可能となる。筆者が普段扱っている機材を例にとると、GoPro Hero12は最小焦点距離が12mmとされており、約156度の非常に広い視野角を持つ。一方、Sony HDR-CX680の最小焦点距離は26.8mmで、正確な視野角は公表されていないが、GoProに比べると画角は明らかに狭いと考えられる。

2.3 実際の録画

■録画する対象 ビデオカメラで会話場面を録画する際には、何を撮影の対象とするかをあらかじめ決めておく必要がある。例えば、全体のやりとりや話者間の相互作用を把握することを重視する場合には、全ての話者の活動が一度に写るよう、俯瞰的なアングルで撮影する方法がある。一方で、話者が座っている場面などでは、顔の表情や視線、手の動きといった上半身の動作に焦点を当てて撮影することも考えられる。また、収録者にとって会話場面や対象となる活動が馴染みのないものである場合には、撮影を通じて観察を行いながら、徐々に撮影の主対象を明確にしていくという方法もある。このようなアプローチは、「撮りながら観察する」（南出・秋谷 2013: 51）という考え方に基づくものであり、探索的な収録において有効な手法とされる。このように、撮影の主たる対象は、研究の目的や分析の視点、対象とする会話の種類などによって異なるため、収録前にこれらを明確にしておくことが重要である。

■俯瞰的に会話を録画する場合 会話場面を俯瞰的に撮影する場合、カメラを話者から一定の距離をとった位置に設置し、複数の話者を同一画面内に収めるのが基本である。距離が確保できる場合は通常のレンズでも話者全員を画角に

収めることができるが、撮影場所が狭い場合や設置に制限がある場合は、広い視野角のビデオカメラや広角レンズを使うことで対応できる。

図5に、異なる視野角のカメラで同じ場所を撮影した画像の例を示す。



図5　さまざまな視野角での撮影範囲（上：デジタルビデオカメラ、中：広角レンズ、下：ウェアラブルカメラ）

上段は一般的なデジタルビデオカメラ（Sony HDR-CX470）、中段は同カメラに0.7倍広角レンズ（RAYNOX DCR-732）を装着した場合、下段は広視野角のウェアラブルカメラ（GoPro Hero12）による撮影の例である。視野角が広くなるほど撮影範囲は広がるが、被写体が小さく遠く見え、周辺部に樽型歪みが生じやすくなることが分かる。

■喫茶店などで小さなテーブルを囲んだ会話を録画する場合　喫茶店のような限られた空間で、小さなテーブルを囲んで会話を行う場面では、カメラの設置に制約が生じやすい。特に、テーブルの周囲に三脚を設置する余裕がない場合や、周囲の通行人や環境への配慮が求められる場面では、小型カメラを使った省スペースな撮影が有効である。例えば、テーブルの隅に小型のウェアラブルカメラを設置すれば、話者の自然な振る舞いを妨げず、上半身を中心に記録できる。人数が多い場合には、広角レンズ搭載のカメラを使うことで、近距離であっても複数の話者を画角内に収めやすくなる。テーブルの端に機器を置けない場合は、会話の場の中央にカメラを置く方法もある。例えば360度撮影可能なKodak PIXPRO SP360 4K（以下、SP360）のようなカメラを使えば、1台でテーブルを囲む全ての話者を同時に撮影できる。

CEJCでは、GoPro Hero3+を2台、SP360を1台、合計3台のカメラで撮影する方法を基本としている。図6は、飲食店における5人の会話を撮影した際の映像の例である。GoProをテーブルの両端に設置し、SP360を中央付近に配置することで、5名の話者を記録することができている。



図6　喫茶店での5人会話のビデオ撮影の例（CEJC会話 ID：C001_001）

なお、喫茶店など公共性の高い場所での撮影では、隣席の客や店舗スタッフが大きく写り込まないよう、カメラの向きや画角に注意を払うことが重要である。また小型カメラの普及により目立たない撮影は可能だが、機器が認識されにくいと隠し撮りと見なされる可能性があり、倫理的な問題が生じるおそれがある。たとえ小型のカメラを使用する場合であっても、機器の存在を把握できるように設置する、あるいは了承を得るなどの配慮が不可欠である。

■屋外を移動しながら行われる会話を録画する場合　散歩など屋外を移動しながら行われる会話を収録する際は、固定カメラを用いた撮影は難しく、話者の全身や顔を動画に収めることも困難なことが多い。このような状況では、話者の身体に小型のウェアラブルカメラを装着し、話者の視界や周囲の環境を記録する方法が有効である。CEJCにおいても移動中の会話の撮影にはこの方法を採用した。ウェアラブルカメラを胸部や腕、頭部などに装着することで、話者が見ている風景や立ち止まった場所などをそのまま動画として記録できる。これにより、会話が行われている環境を可視化でき、発話内容とその場の状況との関係を分析するための手がかりとなる。

□3　メタ情報の収集

■会話状況の記録　図7は、CEJCの収録時に用いた会話状況の**記録ノート**を簡略化したものである（田中ほか 2018）。

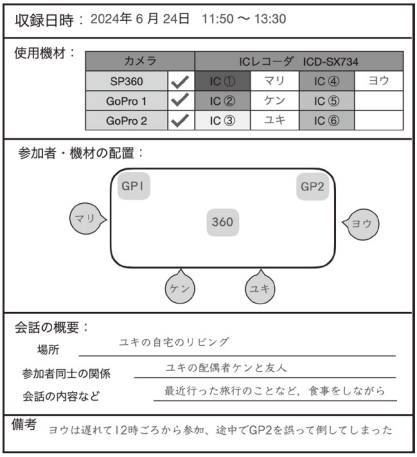

図7　記録ノートの例

記入項目は、収録日時や使用した機器、会話に参加した話者、機器や話者の位置関係、会話の概要、備考情報などである。CEJCでは収録に利用する機器が決まっているため、使用機器にチェックを入れる書式となっている。またCEJCでは、機器の設定は固定だが、収録ごとに変更する場合には設定を記録する欄を設けた方がよい。CEJCでは原則として話者全員が個別のICレコーダーを身に付けて録音したため、ICレコーダーに番号を付与し、誰がどのレコーダーを使ったかが分かるような書式になっている。また収録時の話者と機器の位置関係や、収録状況や機器トラブル等を記録しておくと、あとで収録データを整理する際に参考になる。

■話者情報の収集　話者に関する情報も必要な範囲で収集する。例えばCEJCでは、年齢層（5歳刻み[1]）、性別、出身地、現在の居住地、居住歴、職業などの情報を収集している。会話の場合、話者間の関係性の情報を記録することもある。第1章3節で言及したように、CEJCの大半を占める**個人密着法**による収録では、40名の協力者に会話の収録を依頼したことから、協力者から見た関係性の情報を体系的に収集した。一方、協力者以外の話者間の関係性については体系的に収集しなかった。大人数の会話も少なからずあり、全ての組合せの関係を調べることが難しかったためである。ただし、後日行った協力者に対するヒアリングなどから把握できた範囲の情報を備考情報として提供している。また親疎関係や先輩・後輩などの上下関係に関する情報も体系的には収集せず、把握できた範囲の情報を備考で触れるに留めている。

これらの情報を記入する書式には、年齢や出身地など個人情報を含むため、氏名を書く欄は設けずに話者ごとに付与したIDのみ記録し、IDと氏名・連絡先等の対応表は別に作成して管理する。

□4　収録後の音声・動画データの加工

本節では、音声・動画データの収録後の工程として、収録データの整備や、コーパスとしての公開に必要となる音声・動画データの加工の基本について説明する。なお、2025年現在、入手可能であり、主要なプラットフォーム（Windows, Mac, Linux）において無料で利用可能な初心者向けのツールを用いた方法を紹介する。

1| 生年月日を記入してもらい、収録日と生年月日から年齢層を求めた。CEJCでは数ヶ月にわたって会話を収録したため、収録時期により年齢が変わる可能性があったので、生年月日を記載してもらったが、こうした事情がない限り個人情報を過剰に収集すべきではない。

4.1　音声・動画の形式変換

収録後に、ファイルサイズを小さくするなどの目的でファイル形式を変換したり、動画データから音声を抽出したりすることがある。このような処理を簡便に行うことができる無料のソフトウェアの1つに**VLC media player**がある。VLC media playerでの形式変換の手順はWindowsとMacで異なる。

Windowsの手順：

1. 上部メニューバーから「メディア」→「変換／保存（C）」を選択。
2. 「ファイル」タブ→「追加」で変換対象の動画・音声ファイルを選択。
3. 「変換/保存」をクリックすると「変換」ウィンドウが表示（図8左）。

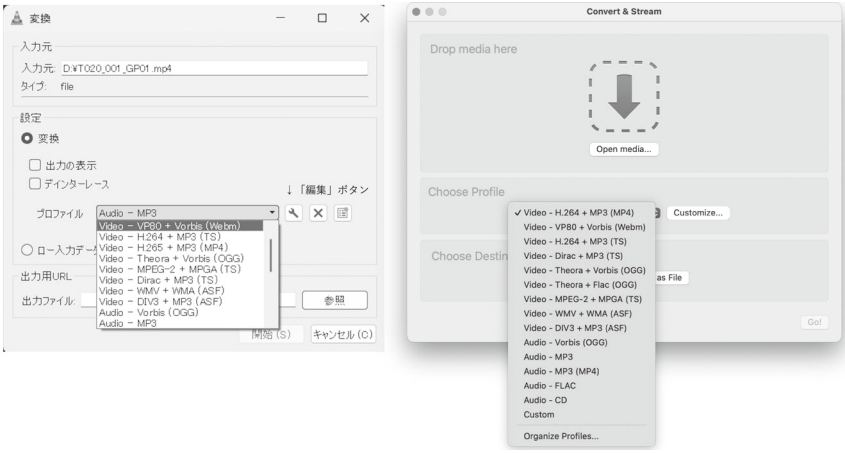

図8　VLC media playerでの形式変換（左：Windows、右：Mac）

4. 「プロファイル」のプルダウンメニューから出力形式を選択。右横の「編集」ボタンをクリックすると細かな設定を行うことができる。
5. 「出力ファイル」欄の「参照...」で保存先フォルダとファイル名（拡張子込み）を指定し、「開始」をクリックして変換後のデータを保存。

Macの手順：

1. 上部メニューバーから「ファイル」→「変換／ストリーミング」を選択すると「変換とストリーミング」ウィンドウが開く（図8右）。
2. 「メディアを開く」で変換対象とするファイルを選択。
3. 「プロファイル」のプルダウンメニューから出力ファイルの形式を選んで

変換。「カスタマイズ」を選択することでより細かな設定を行うこともできる。保存方法はWindowsと同じ。

いずれの方法でも、入力を動画データとし、出力形式としてmp3などの音声ファイル形式を選択すると、動画データから音声を抽出することができる。

4.2　音声・動画データの同期

複数の音声レコーダーやビデオカメラで音声や動画を収録する場合、必ずしも収録の開始・終了位置が揃っているわけではないため、複数の音声・動画データを同期する必要がある。図9に同期作業の流れを示す。

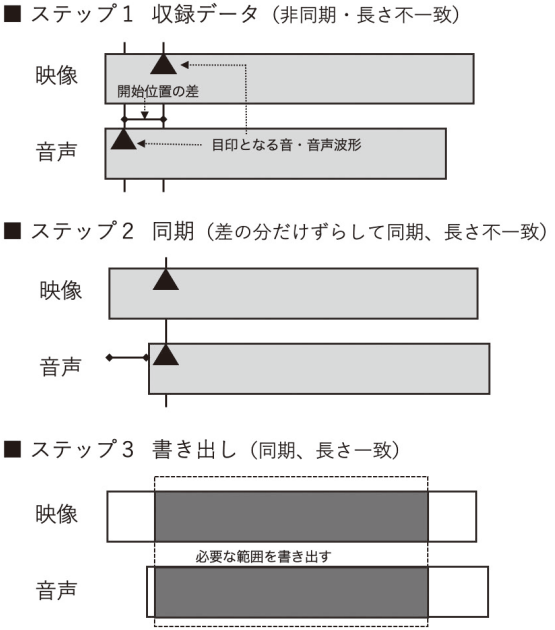

図9　同期作業の流れ

収録した段階では、開始位置は揃っておらず、またファイルの長さも同じではない（ステップ1）。そこで、目印となる音や音声波形を基準に開始位置の差の長さを求めた上で相対的に位置をずらすことで同期させる（ステップ2）。この段階ではファイルの絶対的位置は合っておらずファイルの長さも異なるが、後述する通りこの方法で分析に利用することもできる。データとして必要な範

囲を特定してファイルを書き出すと、データの利便性は上がる（ステップ3）。こうした同期作業を行う方法として、音声や波形などを頼りに手動で行う方法と、同期機能を有するソフトウェアを用いて自動で行う方法がある。

■同期の目印となる位置を手動で探す　まず、音声や音声波形（動画の場合は動画に含まれる音声・音声波形）から、同期の目印となる位置を探す。その際、聴覚に頼った方法だけでなく、視覚的に確認する方法も採用するとよい。波形の確認には**Audacity**や**Praat**などが利用できる。ここではPraatを使う方法を紹介する。なお、Praatの使い方や機能の詳細については、第7章のほか、北原ほか（2017）を参照されたい。

図10上は、「ほい 始めます」と発話し、収録開始の合図として手を叩いた場面の音声波形とスペクトログラムである（ICレコーダーで録音した音声）。

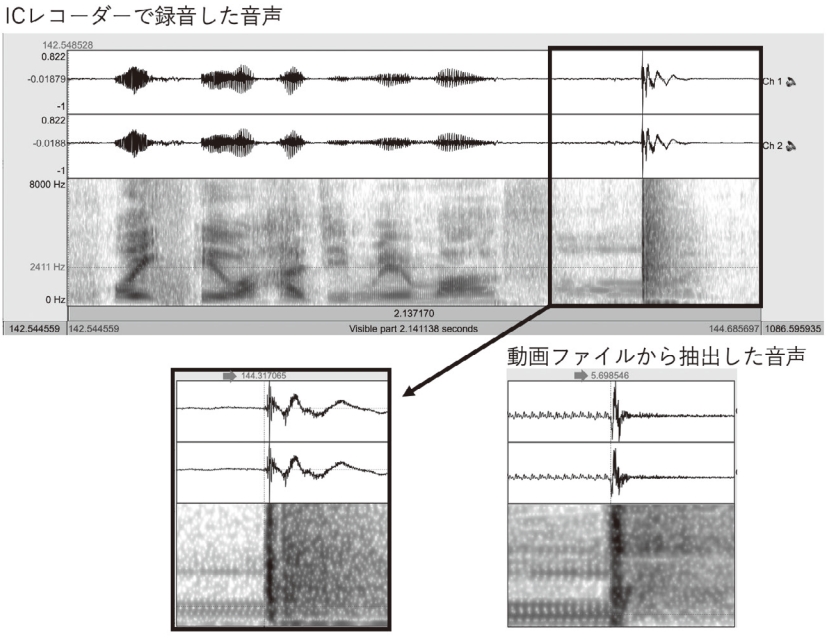

図10　Praatでの目印の時刻情報の確認

手を叩いた音を拡大して音の開始位置を確認すると（図10左下）、144.3秒付近であることが分かる。また、ビデオカメラで録画した動画ファイルから、4.1

節で示した方法で抽出した音声ファイルを対象に同じ箇所の開始位置を確認すると、5.7秒付近であることが分かる（図10右下）。つまりICレコーダーの音声とビデオカメラの動画中の音声の開始位置の差は138.6秒であり、ICレコーダーの方がビデオカメラよりも138.6秒ほど早く録音を開始していたことになる。なお、目印として利用するために、カチンコなど硬いものを打ち合わせた音や手を叩いた音など、鋭い音を収録の冒頭に入れておくとよい。

■手動で同期　ELANを活用し、オフセットの設定をすることで、図9のステップ3の書き出しをせずに同期させることができる。なおELANの使い方や機能の詳細については第6章のほか、細馬・菊地（2019）を参照されたい。ELANでのオフセットの設定は、「編集」→「リンクファイル」から行う。図10で対象とした音声ファイルと動画ファイルの場合、両者の開始位置の差は138.6秒（138600ミリ秒）であり音声の方が早く録音を開始したことから、図11に示す通り、音声ファイルのオフセット値を「138600」（ミリ秒単位）と指定する。これは図9のステップ2に相当する方法で、差の分だけずらして同期させている。



図11　ELANでのオフセットの設定

■自動で同期　本項ではFilmoraを用いて自動で同期させる方法について解説

する。本製品は執筆時点で無料で利用可能[2]だが、製品版の購入もできる。直感的に分かりやすいアプリケーションと思われるため、自動同期を実施したことがない人は無料版で体験してみるのもよいだろう。なお、自動同期の方法や手順は用いるツールに依存するため、自動同期に関わる操作の例として参照されたい。同期の手順は次の通りである（図12）。



図12　Filmoraで複数メディアを自動同期（同期後の画面）

1. アプリケーション起動後の「プロジェクト作成」→「新しいプロジェクト」でプロジェクトを作成。
2. メディアパネル（①）で編集対象とする動画・音声ファイルをインポート。
3. 同期対象とするメディアパネル上の動画・音声ファイルを、下段のトラックパネル／タイムライン（②）上に1つずつドラッグすると、動画・音声トラックがタイムライン上に時系列順に表示される。この図では2つの動画と1つの音声をタイムライン上に配置している（3つ目の動画が薄く見えているが、これについては後述する）。以上、プロジェクトの作成、メディアのインポート、タイムラインへの配置の手順は、後述のFilmoraの各種操作で共通する。

2｜ 無料版では動画に大きく社名と製品名を示す透過性の帯が入る。



図13　Filmoraで複数メディアを自動同期：トラックパネル上の動画・音声トラックを拡大表示

4. タイムライン上で同期させたい全てのメディアトラックを選択し、右クリックで「自動同期」を選択すると自動で同期が開始する。
5. 次の手順で音声や動画を視聴して同期がとれているかを確認する。
   (a) 再生ヘッド（④）を視聴開始位置に移動し、右上のプレビューウィンドウ（③）で再生する。複数の動画をタイムライン上に配置すると、プレビューウィンドウでは重なって表示される。ウィンドウ上の動画をクリックするとサイズや位置を調整できる。③はこの方法で調整している。
   (b) 音声で同期を確認する場合、タイムラインの表示を拡大した上で音声・動画トラック中の音声波形（緑で表示）を比較するか、音を聞いて確認する。複数のメディアを同期する場合、⑤をクリックして動画・音声をミュートにし、数をしぼって視聴すると音声での同期確認がしやすくなる。
6. 微調整が必要な場合、タイムラインの表示を拡大した上で、移動対象となるトラックを左右にドラッグして手動で位置を調整する。

自動同期について2つ補足する。（動画中の）音声を指標に同期させるため、ノイズが多かったり音量が極端に小さかったりすると同期がうまく行かないことがある。またFilmoraの自動同期では、タイムライン上でビデオ1に配置した動画が固定されて基準となり、ビデオ2以降の動画や音声が相対的に移動する。

そのため、ビデオ1には最も早く収録を開始したもの（動かさなくてよいもの）を配置する必要がある[3]。

■範囲を指定して書き出し　図9ステップ3に相当する書き出し手順を説明する。図14はFilmoraで書き出し範囲を指定する方法である。

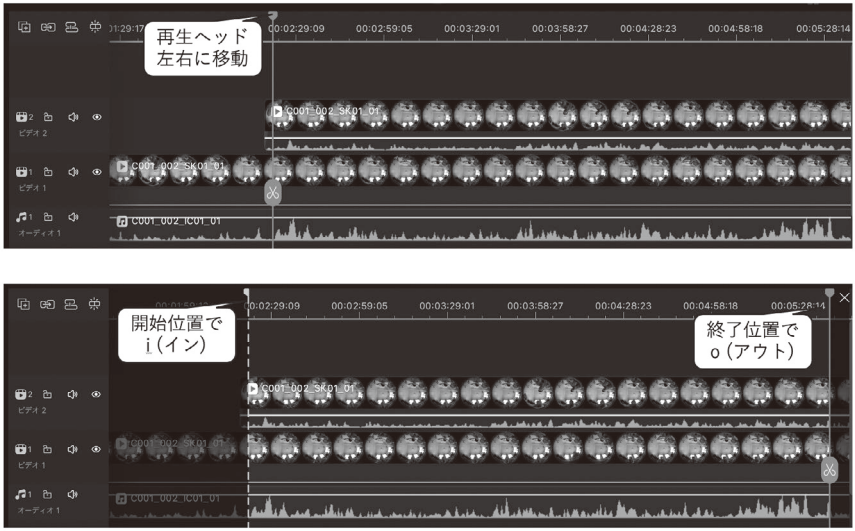

図14　Filmoraで書き出し範囲を指定

図14上段に示すように、再生ヘッドを左右に移動させ、書き出し範囲の開始位置でiキー（イン）を、終了位置でoキー（アウト）を押して範囲を指定する。その上で、書き出し対象とするタイムライン上のトラックを選択し、右クリックで「選択した範囲をエクスポート」を選択して書き出す。なお脚注2でも補足したが、無料版では動画に大きく社名と製品名を示す透過性の帯が入る点に注意する必要がある。

音声ファイルの場合、PraatやAudacityで処理することもできる。ここではAudacityでの手順を説明する。

3｜図13に示したデータは、2つの動画よりも先に音声の録音を開始しているため、ビデオ1を基準とする限り同期はとれない。音声の始端を削除してから同期させる方法もあるが、この例では、ビデオ1と同じ動画をビデオ3にも配置し、ビデオ2、3、及び音声を選択して同期する方法を採用している。ビデオ1はダミーのため⑤でミュートにしている。

1. 上部メニューの「ファイル」→「開く」で音声ファイルを指定すると、図15のように波形が表示される。必要に応じて拡大・縮小する（①）。
2. 波形上でドラッグして削除対象とする範囲を選択し（②）、③で再生して確認する。選択範囲の時刻情報や現在カーソルが置かれている位置の時刻情報は④で確認できる。
3. 「編集」→「削除」で不要な範囲を削除する。
4. 「ファイル」→「オーディオをエクスポート」でファイルを保存する。



図15　Audacityで範囲を指定

4.3　非公開情報の処理

会話の中で話者の名前が発話されたり、名札が動画に写り込んだりした場合、データを公開する前に、当該箇所の音声をノイズで置換したり動画を**モザイク処理**したりするなどの**マスキング加工**を施す必要が生じる。本節では音声と動画に分けてその方法について説明する。

■音声のマスキング　ここではAudacityを使った方法について説明する。手順は次の通りである（図16）。

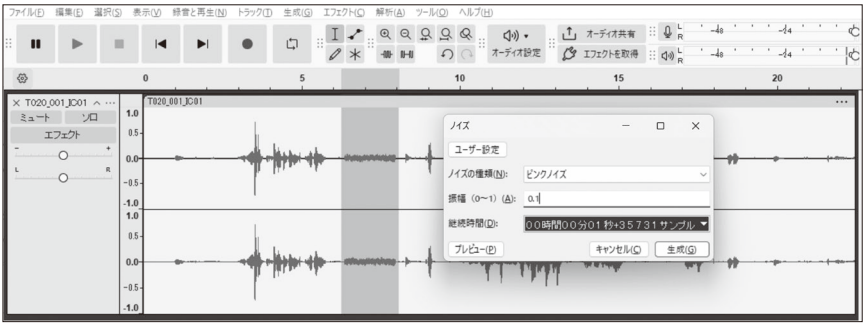

図16　Audacityでのノイズの挿入

1. 波形上でドラッグして範囲を選択し、聴取してマスキング範囲を確定。
2. 上部メニュー「生成」から「ノイズ」を選択し、ノイズの種類や振幅（音の大きさ）を設定して「生成」をクリックする。

ビデオカメラで録画した動画データについても音声が含まれるため、同様にマスキング処理を施す必要がある。方法としては、(1) 動画から音声を抜き出し (4.1節参照)、上記方法でノイズ音に置換した上で、動画の音声トラックをノイズ処理後の音声に差し替える、あるいは (2) 音声レコーダーで収録した音声をノイズ処理した上で、動画ファイルの音声トラックを差し替える方法などがある。いずれの方法においても動画トラックの音声を差し替える必要がある。Filmoraでの手順は次の通りである (図17)。

1. 動画ファイルをタイムラインにドラッグする。
2. タイムライン上の動画トラックを右クリックして「オーディオを分離する」を選択すると、音声トラックが別途タイムラインに表示される。
3. 分離した音声トラックを削除 (右クリックのメニューで削除選択)。
4. ノイズ処理済み音声ファイルをタイムラインにドラッグ。この状態でエクスポートすれば音声を差し替えた動画が生成される。

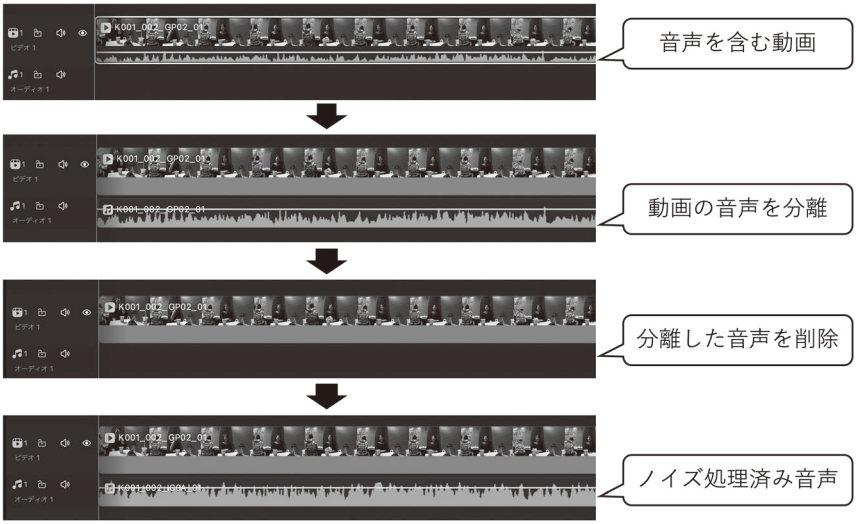

図17　Filmoraでの動画中の音声の差し替え

■動画のマスキング：固定した箇所をモザイク処理する場合　本項と次項では、Filmoraを用いた動画のマスキング加工について解説する。

まず、固定された箇所を対象にモザイク処理を行う方法について説明する。壁の掲示物など対象が全く動かない、あるいはほとんど動かない場合などに用いる。手順は次の通りである（図18）。

1. 対象とする動画をタイムラインにドラッグする。
2. 上部のアイコンから「エフェクト」を選び、左上のパネルの「動画エフェクト」→「ユーティリティ」→「モザイク」(①) を選択してタイムラインの動画の上段にドラッグすると、②のようにモザイクがタイムラインに表示される。
3. タイムライン上のモザイクの帯の左端あるいは右端をクリックして、モザイク処理が必要な時間幅に調整する。②は幅を広げた状態である。
4. タイムライン上のモザイクの帯をダブルクリックすると、上部のプレビューウィンドウにモザイク処理する範囲を示す矩形（四角形）が表示されるため（③）、位置とサイズを調整する。
5. タイムライン上のモザイクの帯をクリックすると右に示される「エフェクト」のパネル（④）から、必要に応じてモザイクの種類やぼかし具合、不透



図18　Filmoraで固定した箇所にモザイク処理をする場合

明度を設定する。

6. 複数の箇所にモザイク処理を施す場合、2〜4の処理を繰り返す。その場合、タイムライン上に2つ以上のモザイクの帯が表示される。

モザイクの結果は上部のプレビューウィンドウで確認することができる。またファイルをエクスポートすれば指定箇所にモザイクが施された動画が出力される。なお、Filmoraでは楕円など矩形以外の形でモザイク処理を行うことはできない。より高度な処理を行う場合には、Adobe Premiere ProやDaVinci Resolveなどの編集ソフトを用いる必要がある。

■動画のマスキング：動く対象を自動で追跡してモザイク処理する場合　人物の顔などモザイク対象が動く場合、それを自動追跡（トラッキング）できる場合がある。手順は次の通りである（図19）。



図19　Filmoraでの動く対象を自動追跡してモザイク処理する場合

1. タイムライン上でモザイクが必要な区間を特定し、その範囲をシーンとして切り出す。手順としては、タイムライン上の再生ヘッドを左右にドラッグし、シーンの始点に移動させた上で、再生ヘッドにあるハサミのアイコン（①）をクリックして分割する。終点も同様に分割する。ファイル全体を追跡する場合、この操作は不要だが、長いファイルの場合には分割してから追跡した方がよい。
2. 切り出したシーンをクリックして選択し、再生ヘッドを追跡シーンの冒頭に移動させた上で、右パネルの「ビデオ」タブ（図17）から「AIツール」を選んで「モーショントラッキング」をオンにする（②）。
3. プレビューウィンドウに現れる矩形を追跡対象に合わせる（③）。
4. 右パネルのAIツール画面で「クリックしてモーショントラッキングを開始」をクリックすると（④）、当シーン内での対象物の動きを追跡する処理が始まる。
5. 左上のビデオで自動追跡結果を確認し、追跡の範囲から外れてモザイク処理が適切になされていない箇所があれば、ビデオ上で手動で修正する。

6. トラッキングが終了すると、右パネルの「モーショントラッキング」の下に「リンク要素」(⑤) のメニューが表示される。そこから「モザイクを追加」を選択し、必要に応じてモザイクの種類などを設定する。するとタイムライン上にモザイクの帯が生成される。

固定のモザイクを施す場合とは異なり、同時に複数の箇所を追跡してモザイク処理を施す方法は手順が複雑である[4]。簡単なのは、自動追跡を繰り返す方法である。具体的には、上記手順に従いモザイク処理を施し、動画ファイルとしてエクスポートした上で、その動画ファイルを再度読み込んで別の対象のモザイク処理を行う、という作業を繰り返す。この方法では、以前の作業で付与したモザイクを編集することはできないことに注意する必要がある。

自動追跡は大幅に作業量を減じてくれるが、それでも手作業での調整は必要である。例えば追跡対象が一時的に部屋から出て行って戻ってくるなどした時には、追跡は途中で途切れてしまうことが多い。また、追跡対象が手前と奥を移動して大きさが変わっても、モザイクのサイズ (追跡対象の範囲) は自動では変更されない場合がある。このような場合には、自動追跡が成功する範囲ごとにシーンを区切り、トラッキングをやり直すなどの対応が必要となる。

□5　おわりに

本章では、会話収録に関する基本的な知識から具体的な収録法、収録後の編集の方法を概説した。紹介した内容は、あくまで会話収録の基本的な枠組みであり、実際の現場では状況に応じた柔軟な対応が求められる。読者が自らの研究目的に応じて工夫を重ね、よりよい収録環境やデータ整備を追求していくきっかけとなれば幸いである。

参考文献

細馬宏通・菊地浩平 (編) (2019)『ELAN入門―言語学・行動学からメディア研究まで―』ひつじ書房.
北原真冬・田嶋圭一・田中邦佳 (2017)『音声学を学ぶ人のためのPraat入門』ひつじ書房.
南出和余・秋谷直矩 (2013)『フィールドワークと映像実践―研究のためのビデオ撮影入門―』ハーベスト社.
田中弥生・柏野和佳子・角田ゆかり・伝康晴・小磯花絵 (2018)「『日本語日常会話コーパ

4| トラッキングとモザイク処理した動画を複製し、タイムライン上に重ねて配置した上で別のトラッキングとモザイク処理をする方法もある。書き出さずに複数のモザイクを同時に行えるが、手順が複雑なのでここでは省略する。

ス』の構築―会話収録法に着目して―」『国立国語研究所論集』14: 275–292. http://dx.doi.org/10.15084/00001424

関連URL（2025年7月31日確認）

Audacity
https://www.audacityteam.org/
CEJCメタ情報
https://www2.ninjal.ac.jp/conversation/cejc/metainfo.html
ELAN
https://archive.mpi.nl/tla/elan/
FFmpeg
https://ffmpeg.org/
Filmora
https://filmora.wondershare.jp/
Praat
https://www.fon.hum.uva.nl/praat/
VLC Media Player
https://www.videolan.org/index.ja.html

（臼田泰如）