取調室の言動の解析と可視化

代表研究者	豊　浦　正　広	山梨大学　大学院総合研究部工学域　教授
共同研究者	西　﨑　博　光	山梨大学　大学院総合研究部工学域　教授
共同研究者	西　口　敏　司	大阪工業大学　情報科学部　教授

1 はじめに

密室で行われる警察・検察による取調では、被疑者は法律的な前提知識を持たない不利な立場に置かれ、自白を強要されるなどの問題が繰り返し報告されている。袴田事件(2024 年再審無罪確定)、Winny 事件(2011 年無罪確定)、厚生労働省元局長事件（2010 年無罪確定）など、取調方法そのものが社会問題化している。2016 年の刑事訴訟法改正により、裁判員裁判対象事件・検察官独自捜査事件については録音・録画が義務付けられたが、これらは身体拘束事件全体の 3〜4%に過ぎない。日本弁護士連合会は全事件の取調の可視化・立会いを求めて国への働きかけを続けている。

しかし、全事件の取調が録音・録画されたとしても、そのすべてを弁護士が確認することは現実的ではない。立会いが認められても、その場で問題となる言動を制止することは困難である。従来の「可視化」とは、長時間の映像や録音を人間が目と耳で追うことを指すが、情報通信技術の観点からは、情報が集約され短時間で効率的に一覧できる状態こそが真の可視化である。本研究は、取調室の録画・録音およびセンサデータから問題となる言動を自動検出し、必要に応じてその場で警告することを目指す。機械学習、大規模言語モデル（LLM)、IoT センシング技術を統合し、取調の公正性を客観的に評価可能な形で可視化する基盤技術の確立を目標とした。

研究は山梨県弁護士会弁護士ほかの監修のもと、次の 4 項目を柱として 2025 年 4 月から 2026 年 3 月まで進めた。(1) 音声からの問題発言の検出、(2) 映像・振動センサからの問題行動の検出、(3) 取調を妨げない問題行動の警告提示、(4) 長時間の映像・音声の可視化である。

本研究の着想は、取調側の「拒否感」をいかに軽減するかという問題意識にも基づいている。正当な取調を妨げる装置は採用されない。映像や録音はその場にのみ保存され、解析のためにサーバへデータを送信しない設計が望まれる。プライバシーへの配慮も不可欠であり、取調官を傷つける用途であってはならない。申請者らは研究機関開始時までに、イスの振動から座席者を高精度で識別する匿名センシング技術を開発してきた。この知見を取調室にも適用し、双方に拒否感を与えず、かつ公正な取調を支援するシステムの実現を目指した。共同研究者の西崎博光は大規模言語モデルと RAG（Retrieval-Augmented Generation）による専門用語認識の精度向上、西口敏司は映像解析と警告提示装置の設計を担当し、3 名体制で研究開発を進めた。

情報通信技術の急速な発展、とりわけ ChatGPT 等の大規模言語モデルの一般化により、機械が人間の言動の文脈を理解できる段階に至っている。本研究課題を通じてこの技術の恩恵を刑事司法の現場に還元することは、情報通信技術の社会的応用として意義が大きい。石黒らの「対話知能システム学」や西田らの「会話情報学」など、対話のメカニズムを理解する学術的蓄積は豊富であるが、取調室における問題言動の検出に特段の焦点を当てた研究は十分ではない。法社会学における会話分析の知見も参考にしつつ、リーガルテックの文脈で実装可能な技術開発を目指した。

2 研究内容と成果

2-1 音声からの問題発言の検出

（１）テキストベース検出手法の開発

音声から問題発言を検出する第一段階として、OpenAI Whisper による自動文字起こしと、GPT-4o 等の大規模言語モデルによる文面解析を組み合わせた手法を開発した。LLM に対しては、威圧的な発言、不適切な皮肉、被疑者の理解を超える専門用語の使用、一方的な長時間の説教、相手の発言を遮る言動などを検出対象として明示したプロンプトを設計し、反復的に改良を加えた。

公益財団法人電気通信普及財団
研究調査助成報告書　第41号　2026年度

データセットとしては、公開されている不適切な取調音声を中心に検討を進めた。模擬取調の撮影も検討したが、専門知識を持たない研究室メンバーのみでは再現性の確保が困難であり、既存の実録音源を用いた方法論の確立を優先した。実際の取調室への導入および現場データの入手は期間内には困難であったが、公開音源と研究室環境での収録データにより、手法の有効性と限界を検証した。

検証の結果、テキスト化された発話内容からは、明確な威圧表現や不適切な命令調の発言をある程度自動検出できることが確認された。一方で、方言、強い掛け声、抑揚や間の取り方など、音声そのものに含まれる非言語情報は文字起こしだけでは判別が難しいという課題が明らかになった。この知見は、後述する音響特徴ベース検出の併用を研究方針として位置づける契機となった。

Whisper による文字起こしでは、json 形式でセグメント単位のタイムスタンプを取得し、後段の可視化システムと連携可能なデータ形式を整備した。LLM への入力では、取調の文脈を考慮したシステムプロンプトを設計し、検出結果には該当発言の引用と問題性の理由を添える形式とした。申請書で計画していた RAG 構築については、法律用語や特定分野の事件で頻出する語句を参照データベースに登録し、内容認識の精度向上を図る基盤の設計を行った。完全な現場適用に至るまでにはさらなるデータ拡充が必要であるが、公開音源を用いた検証により、手法の基本構成の有効性は確認できた。

共同研究者西崎の担当領域である音声認識と LLM の連携では、法律・捜査分野に特有の語彙が一般の音声認識モデルでは誤認識されやすい点も確認した。RAG により判例用語や訴訟実務上の固定表現を参照させることで、後処理の LLM 解析の精度を補完する設計とした。

（２）テキストベース検出手法の開発

テキストベース検出の限界を補うため、発話内容に依存しない音響特徴量の解析手法を設計した。具体的には、音量レベルの時間変化、発話比率（話者間の発話時間配分）、一方的な長時間独話の検出、発話の遮断（オーバーラップパターン）、応答潜時（質問終了から回答開始までの沈黙時間）などを指標として抽出する方針とした。人間の声は心理的ストレス下でピッチ、音量、話速などに特徴的な変化が生じることが知られており、取調官の声の急激な高まりや被疑者の声の不安定さは、不適切な圧力の存在を示唆する可能性がある。

テキストベース検出と音響特徴ベース検出を統合し、まず音響特徴で「疑わしい」時間帯を絞り込み、該当区間のみ LLM による詳細な文面解析を行う二段構成の手法を設計した。これにより、長時間の取調記録に対する計算コストの削減と、検出精度の向上の両立を目指した。

音響特徴の抽出では、フレーム単位の RMS エネルギー、ゼロ交差率、スペクトル重心などの低レベル特徴量に加え、発話区間の検出と話者間の発話時間比率を算出する処理を組み込んだ。取調官の発話占有率が極端に高い場合は被疑者の発言機会が制限されている可能性があり、異常に短い応答時間は誘導や圧迫による即答、長すぎる沈黙は困惑や心理的負荷を示唆する。沈黙は取調の文脈では心理戦術として用いられることがあり、一定秒数以上の沈黙の発生回数とその前後の話者も記録対象とした。これらの指標は、後述の可視化システムにおいて時系列グラフとして表示し、公正性評価の客観的資料とすることを想定している。

（３）会話要約・音声再生システムの開発と現場展開

取調室への直接導入の前段階として、弁護士の日常業務における LLM 活用の有用性を実証するため、依頼人との面談記録を効率的に整理する Web アプリケーション“Audio Scribe & Summarizer”を開発した。図 1 にシステムの概要を示す。本システムは OpenAI Whisper API による音声文字起こし（セグメント単位のタイムスタンプ付き）、GPT-4o による要約（簡潔・中程度・詳細・箇条書き・アクションアイテム・精密要約の 6 段階）、音声プレーヤーと文字起こしの同期表示（文字起こしの該当箇所をクリックすると音声がその位置から再生）をブラウザ上で実現する。複数 MP3 ファイルの一括処理にも対応している。

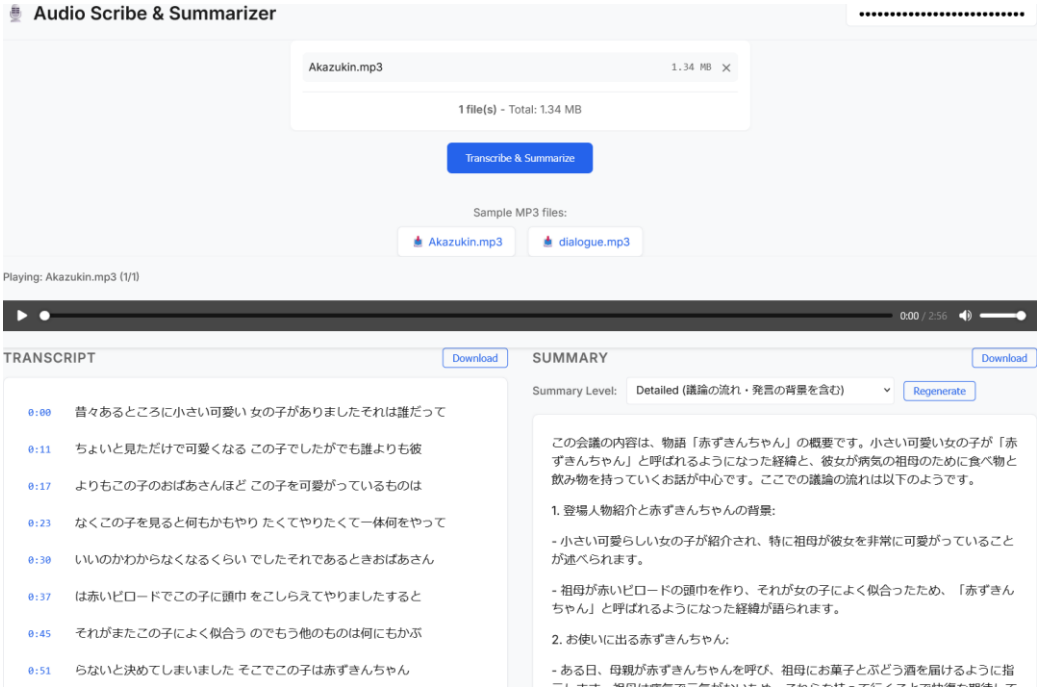

図 1 音声要約 Web アプリケーション Audio Scribe & Summarizer

本システムを山梨県弁護士会所属弁護士ほか 2 名に提供し、依頼人との面談記録の整理・確認に継続して使用してもらっている。長時間の面談音声を聞き返す負担を大幅に軽減し、要点の把握と記録の確認を支援する実用的なツールとして定着しつつある。取調記録の分析にも転用可能な技術基盤であり、弁護士と研究側の信頼関係を構築する上でも重要な成果である。

システムの操作は Web ブラウザ上で完結し、MP3 ファイルをドラッグ＆ドロップでアップロードした後、文字起こしと要約を一括実行する。文字起こし結果はタイムスタンプ付きで表示され、音声プレーヤーの再生位置と連動して該当セグメントがハイライトされる。ユーザーは文字起こしの任意の箇所をクリックすることで、その時点から音声を再生でき、長時間記録の確認効率が向上する。要約レベルは 6 段階から選択可能であり、用途に応じて簡潔なポイント整理から詳細な議事録に近い出力まで切り替えられる。文字起こしテキストと要約結果はそれぞれダウンロード可能である。映像ファイルからの音声抽出については、ブラウザ上での大容量映像処理を避け、ffmpeg による MP3 変換を事前処理として弁護士に案内する運用とした。話者分離については Gemini API や Google Cloud Speech-to-Text も試行したが、時刻のずれや認証の問題などがあり、現時点では Whisper 単体の文字起こしを主軸としている。今後、Whisper と pyannote 等を組み合わせた話者分離の実装を検討している。

2-2 映像・振動センサからの問題行動の検出

（１）映像による問題行動検出の検討と限界

映像からの人間の活動推定については、MediaPipe Pose Landmarker など手軽かつ高精度な姿勢推定ツールを用いた検討を行った。しかし、想定する取調室のような小部屋環境では、部屋を俯瞰するカメラの映像からは姿勢や表情の変化を捉えにくく、卓上にカメラを置いても手元や足元が隠れることが多いことが判明した。カメラの設置位置は現場ごとに異なり、訓練データを取得した環境とは異なる条件での推定が必要になることも確認した。

これらの結果から、映像単独での問題行動検出には構造的な限界があると判断し、プライバシー情報を直接取得しないセンサデータとの併用を研究の主軸とした。申請者らがこれまで開発してきた匿名センシング技術の知見を活かし、イスの振動から座席者を 97.7%の精度で識別できる実績を踏まえ、振動センサ中心のアプローチを推進した。取調室では被疑者と取調官が向かい合って着席するため、各自のイスや机にセンサを配置することで、カメラによる顔画像や姿勢の取得なしに身体的行動の変化を捉えることが可能である。映像解析は補助的な情報源として位置づけ、ジェスチャ推定や姿勢変化の急激な変化検出など、参考研究の

知見は引き続き活用する方針とした。

（２）振動センサとオートエンコーダによる異常検出

体やイスに取り付けた慣性センサ（加速度・角速度）から得られる時系列多次元信号を入力とし、オートエンコーダ（AE）による異常検出アルゴリズムを設計した。AE は正常データのみで訓練され、入力信号を低次元の潜在空間に圧縮して再構成する。正常パターンに近い信号は再構成誤差が小さく、訓練データにない異常パターンは再構成誤差が大きくなる。この入出力差分を異常度スコアとして算出する。

さらに、変分オートエンコーダ（VAE）に Transformer や LSTM のモジュールを組み込んだモデルも検討し、時系列の長期依存関係を捉える構成を設計した。大きな信号（机を叩く、急激な身動きなど）の検出だけでも一定程度の異常検出が可能であることが確認された。1 段目で大きな信号を異常として検出し、2 段目で残りの小さな信号に VAE を適用する二段構成により、大小両方の異常パターンを捉える方針とした。

研究室環境でのミーティング中の映像・音声・振動データの収録を行い、基礎データを整備した。通常のミーティングでは問題行動の再現は困難であるが、演技データを異常サンプルとして加えることで、異常検知精度の検証用データセットとして拡充する方向で作業を進めた。

VAE の訓練は正常動作データのみを用い、テスト時に正常データを入力した場合は再構成誤差が小さく、異常データを入力した場合は潜在空間での表現が正常パターンから乖離し再構成誤差が大きくなることを利用する。Transformer モジュールは時系列の長距離依存関係のモデル化に、LSTM モジュールは順序情報の保持に寄与する。エレベータ内の犯罪検出などに用いられる異常検知の知見と、申請者らの匿名センシング研究で培った振動信号解析の経験を統合したアプローチである。映像から得られる姿勢推定結果と振動センサデータを組み合わせるマルチモーダル統合も設計段階で検討し、VIA 上で複数ソースを同期表示する基盤と連携させる方針とした。

（３）発話時顔部モーションによるユーザ識別

映像解析の基盤技術として、MediaPipe の Face Mesh を用いた顔部ランドマーク抽出と時系列ニューラルネットワークによる動作識別の研究も並行して進めた。仮想空間におけるなりすまし防止には継続的な本人認証が必要であるが、意識的な認証操作はユーザ負担となる。そこで日常会話と同様の短いフレーズを発話した際の顔の動きからユーザを識別する手法を検討した。

ノートパソコン内蔵 Web カメラ（25 fps）で顔画像を取得し、MediaPipe Face Mesh から目・鼻・口・鼻と口の領域のランドマークを抽出した。識別には LSTM と Transformer（Encoder 部）の 2 種類のニューラルネットワークを用い、各フレーズを 30 回発話したモーションデータを収集した（18〜43 歳の男女 41 名、平均年齢 21.6 歳）。顔部位ごとの識別精度を比較した結果、口の動きが最も識別に寄与し、鼻と口の領域を組み合わせることで精度が向上した。複数フレーズをひとつのニューラルネットワークに対して学習させる方式では、Transformer モデルが 3 フレーズ全体で正解率 98.3%のユーザ識別に成功し、単一フレーズのみで学習した場合と比較して識別精度が大幅に向上した（LSTM は正解率 0.062〜0.174 向上、Transformer は 0.031〜0.055 向上）。

MediaPipe による顔部モーションの時系列解析と Transformer によるパターン認識の知見は、映像を補助情報源とする本研究の技術基盤を補強する。口部モーションの取得が困難な条件下では鼻と口の領域のモーションデータで代替可能である点も、カメラ設置条件が制約される取調室環境での実装検討に示唆を与える。

2-3 取調を妨げない問題行動の警告提示

（１）研究課題の設定

取調室に AI エージェント型のモニタリング装置を導入する場合、問題行動を検出した弁護士に警告を伝える必要がある。しかし、警告の受信そのものが取調官や被疑者に悟られれば、取調の雰囲気を乱し、正当な取調を妨げる結果になりかねない。Dowell & Berman[1]は、カウンセリング中の非言語的行動（視線の移動など）が共感性や信頼性の評価を低下させることを示しており、Giebelhausen ら[2]は、サービス場面でのデジタル端末への注意がサービス品質の低下を招くことを報告している。取調室においても同様に、警告受信時の身体的反応が対話相手に知覚されない「秘匿性」が警告方式選定の最重要条件となる。そこで、対面会話中の秘匿通知方法の探索と評価を実施した。

対話への介入装置は、取調官と被疑者が背中合わせに向かい合い、それぞれにモニタ・マイク・カメラ・振動センサを備えた構成を想定している。警告は反対側には見えない形で、必要なときにだけ表示される。映像提示は緩やかに行い、モニタを見ていることを相手に気づかれない設計とする。装置の基本構成要素（モニタ、マイク、カメラ、振動センサ、処理用サーバ）は準備済みであり、秘匿通知評価の結果を反映した統合実装が次のステップとなる。

（２）通知方法の探索と選定

Logas ら[3]のスマートグラス通知に関する研究を参考に、通知受信時に身体的反応を伴わず、対話相手に通知受信を悟られない通知方法を「秘匿通知方法」と定義した。選定にあたっては、(a) 対面会話状況における社会的受容性、(b) 技術的成熟度（一般市場で入手可能なデバイスで実現できること）を包含基準とし、9 つの選定観点（過度な視線移動を伴わないか、受信者が確実に通知を受信できるか、通知受信が物理的に相手に露呈しないか、身体的反応を強く誘発しないか、装着時の見た目が不自然でないか、即時性があるか、不快感を与えないか、安全上実装・実験が現実的か、一般的に普及している技術か）により候補を絞り込んだ。

最終的に以下 8 方式を本実験の対象とした。図 2 に実験環境を示す。振動通知：手首、足首、腰（腸骨稜辺り）、ポケット（大腿部辺り）。音通知：イヤホン（左耳）でのビープ音、骨伝導（両耳）でのビープ音。光通知：卓上モニタの点灯、対面モニタの点灯（スマートグラス装着時の通知を模擬）。振動デバイスは XIAO-ESP32-C3 とコイン型振動モータ（uxcell 10mm×3mm, DC 3V）を用いて自作し、PWM 駆動（デューティ比 15.7%）で統一した条件で実験した。その他のデバイスには Shokz OpenMove（骨伝導）、Amazon Echo Buds（イヤホン）、Thinlerain HD-101（卓上モニタ）、I-O DATA EX-LD2702DB（対面モニタ）等の市販品を使用した。

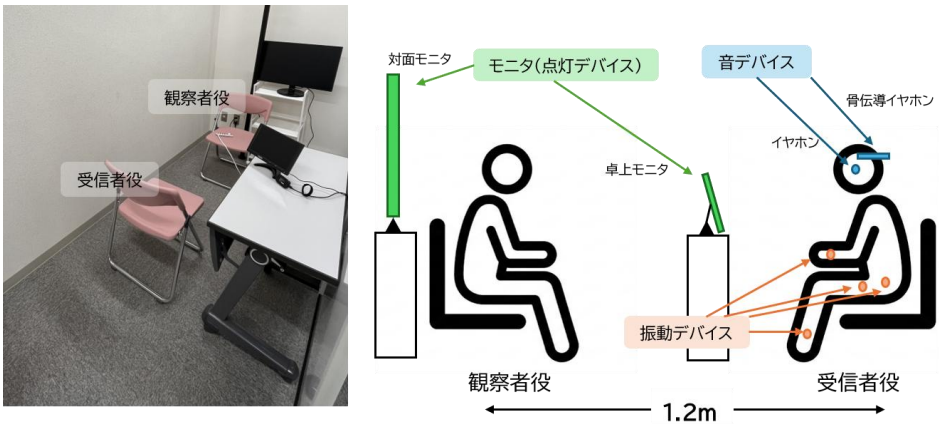

図 2 秘匿通知性検証実験 （山梨大学工学域研究倫理審査委員会承認済(2025-017)）

通知方法の探索は三段階で行った。第一段階で既存の通知方法を洗い出し、社会的受容性と技術的成熟度の観点から机上検討で候補を選定した。第二段階で選定された方式について実験用デバイスを作成し、予備実験（しりとりタスク＋事後評価）を実施した。第三段階で予備実験を通過した 8 方式のみを本実験に採用した。選定観点で×が 3 つ、または×2 つと△1 つ、または×1 つと△2 つがある方式は却下する基準を設けた。卓上モニタの点灯は「過度な視線移動を伴わないか」の観点では不十分と判断されたが、一般的に使用される通知方式としてベースライン比較の対象に残した。

（３）実験設計

実験環境は Hall[4]の対人距離理論に基づき、社会距離と個人距離の境界である 1.2m に設定した。2 名 1 組（通知受信者役・観察者役）が対面着席し、しりとりタスクを行う中で各通知方式による通知を受信した。しりとりを採用した理由は、発言者が交互に入れ替わることで適切な秘匿率を算出できるためである。観察

者は受信者の視線移動や身体的反応から通知受信に気づいた場合にボタンを押下した。被験者は 12 ペア 24 名（男性 23 名、女性 1 名、年齢平均 21.13 歳）、各しりとりタスクは約 346 秒、8 方式×3 回の通知をランダム順で実施した。

評価指標は秘匿率（通知区間において観察者に気づかれなかった割合）と、無通知区間の真陰性率（通知がない区間で誤ってボタンが押されなかった割合）とした。秘匿率が真陰性率を上回る通知方式は「秘匿性がある」と判定する。統計分析には一般化推定方程式（GEE）と Wald 検定（Holm 補正、有意水準 5%）を用いた。

本実験では通知内容をすべて 0・1 の信号に統一し、イヤホン・骨伝導ではブザー音のみ（言語音声なし）、ディスプレイでは画面を光らせるのみ（文字表示なし）とした。これにより「通知の意味理解」ではなく「通知の存在知覚」に焦点を絞った。各デバイスへの通知は 3 回ずつ、計 24 回（8 デバイス×3 通知）をランダム順で送信し、通知間隔は 10 秒±2 秒の乱数、無通知区間 10 秒を 6 回ランダムに挿入した。判定時間は 9.5 秒とし、通知後のボタン押下を通知知覚とみなした。全デバイスを同一実験内で同時装着することで、通知方式間の比較条件を均一化し、観察者が特定方式の反応パターンを学習する問題を回避した。また、観察者が予め通知の存在を知っている状況で実験した。これは秘匿性評価において厳しい条件設定であり、この条件下でも秘匿率が高い方式は実際の業務場面でさらに高い秘匿性を示すと期待できる。

（４）本実験の設計

8 つの通知方式いずれも、対面会話状況において一定の秘匿性が示唆された。ただし方式間には差があり、卓上モニタの点灯は平均秘匿率が最も低く、視線移動を伴う点灯確認が観察者に発覚していた可能性が示された。手首・足首・腰・ポケットへの振動通知は、触覚刺激が無意識の視線移動を誘発しやすく、秘匿率が相対的に低い傾向が見られた。イヤホン（左耳）でのビープ音は骨伝導と比較してやや知覚されやすい結果となった。対面モニタの点灯は Logas らが模擬したスマートグラス通知に相当し、相手の背後にモニタを設置して通知受信を再現した。一方、骨伝導（両耳）でのビープ音通知は、卓上モニタの点灯と比較して観察者に気づかれにくいことが統計的に確認された（Odds Ratio = 2.60, z = 2.74, pHolm = 0.04 < 0.05）。両耳からの同一聴覚刺激は上丘を介して「音源は正中（0 度）」と処理され、すでに正面を向いている場合には視線移動が起こりにくいという神経科学的メカニズムが考察された。

手首への振動通知は秘匿率が低く、触覚刺激が無意識の視線移動を誘発した可能性が指摘された（Amlôt & Walker[5]）。対面着席という「反応が露呈しやすい厳しい条件」で秘匿性が確認された方式は、立位など他の状況でも有効である可能性が高い。今後の課題として、0・1 信号ではなく意味を含む通知内容を対象とした実験、データ数の追加による統計的検出力の向上が挙げられる。

これらの結果は、取調室における弁護士への警告提示には、骨伝導イヤホンによる低音量の音通知が最も適していることを示唆する。対面モニタの隅への視覚的警告も有効な候補である。手首振動など視線移動を誘発しやすい方式は取調室への適用には不向きである。背合わせの 2 モニタと各種センサを備えた介入装置の基本構成要素も準備済みであり、本研究成果を組み込んだ統合システムの実装が今後の課題となる。

観察者向けアンケートでも、卓上モニタの点灯について「相手が手首を見た、きっと振動を受け取ったに違いない」等の回答が得られ、視線移動に基づく通知知覚が確認された。骨伝導通知については「相手に気づかれなかった」という評価が相対的に高かった。Czerwinski ら[6]の通知研究は受信者の作業効率に焦点を当てており、対話相手からの知覚という観点は十分でなかった。本研究は Logas ら[3]の対話相手視点の評価を発展させ、取調室のような権力関係を伴う対面場面への設計指針を提供する点に学術的意義がある。

2-4 長時間の映像・音声の可視化

（１）VIA を活用した統合可視化システム

U. Oxford のグループが開発する VGG Image Annotator（VIA）を基盤とした統合可視化システムを構築した。VIA は映像フレームへの注釈付けに広く用いられるツールであり、本研究ではこれを拡張し、音声・映像・振動センサなど複数のデータソースを時系列で同期表示する体制を整えた。自動検出アルゴリズムが「疑わしい」と判定した箇所をハイライト表示し、検証者が効率的に重点確認できる半自動検証システムも整備した。

従来、数時間に及ぶ取調記録を最初から最後まで映像・音声を追う作業は、弁護士にとって膨大な時間と集中力を要する。本システムにより、検出されたイベントをシークバー上に配置し、必要な箇所だけを効率

的に見返すことが可能となる。検出が不完全であっても、可能性の高い個所を示すことで、人による最終確認の効率を大幅に向上させる設計とした。

VIA への入力形式を整備し、音声・映像・振動センサから得られた検出イベントを時刻情報とともにアノテーションファイルとして出力する処理を実装した。図 3 に概要を示す。検証者は VIA 上でイベント一覧を参照しながら、該当箇所の映像・音声・センサ波形を同時に確認できる。問題行動の検出精度が 100%でなくても、候補箇所の提示により人間の確認作業を支援する「半自動検証」の考え方は、現場データが限られる本研究の条件下において特に重要である。模擬裁判等での利用を想定した実用上の問題点の洗い出しも計画に含まれており、今後のフィードバック収集に向けた基盤が整った。

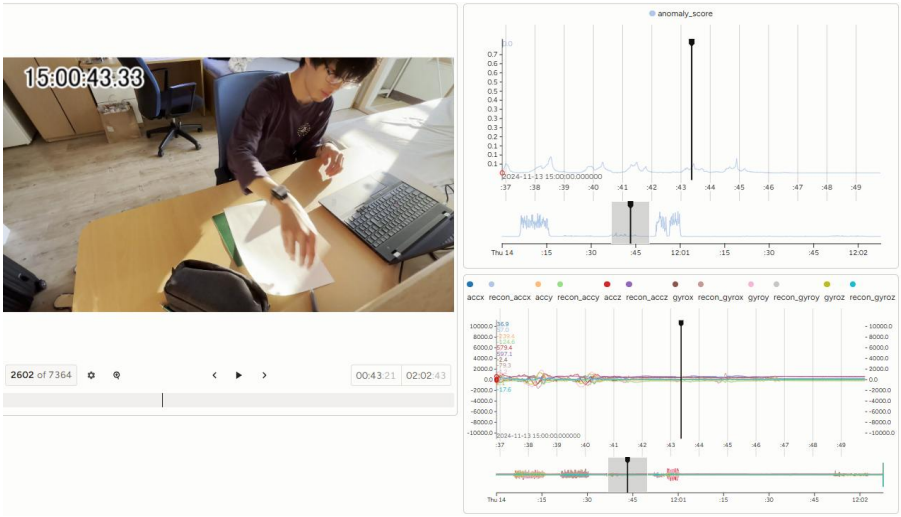

図 3 VIA による音声・映像・振動センサの表示例

共同研究者西口の担当領域である映像解析の知見は、VIA 上でのフレーム単位アノテーションと、姿勢推定結果の時系列表示に活かした。長時間映像の「頭出し」作業を自動化することで、弁護士や研究者が問題箇所の確認に集中できる環境を整備した。

（２）公正性評価指標の設計

取調の公正性を定量的に評価する複合指標の設計も検討した。発話比率と応答時間（会話主導権の指標）、声のトーン・音圧変化（心理的負荷の指標）、沈黙の長さと頻度、身体動作の傾向（非言語的行動の指標）などを統合し、「会話主導権スコア」「心理負荷スコア」等のインジケータを時系列で可視化する方針を整理した。これらはリアルタイムダッシュボードと事後分析レポートの両方に展開可能である。コールセンター分野で通話中の会話時間や応答時間をリアルタイム監視する知見と同様に、取調に特化した指標のリアルタイム表示により、その場での問題検知や振り返りによる改善が期待できる。

2-5 実行過程での問題点と対応

計画時点から想定されていたが、実際の取調室への導入は法的・制度的なハードルが高く、現場データの入手は期間内には困難であった。計画に従い、期間内に基盤技術と基盤システムを完成させ、期間終了後の現場投入を目指す方針とした。オフラインで利用可能なローカル LLM を搭載した機器の準備と弁護士事務所への展開交渉も進め、Audio Scribe & Summarizer による現場連携を実現した。

データセットの準備とテキスト・音響統合手法の設計、VIA 可視化体制の構築、警告提示に加えて、秘匿通知の本実験完了、会話要約システムの弁護士への展開、振動センサ異常検知の詳細設計、研究室データの収録を実施し、具体的成果に結びつけた。実際に動くシステムを弁護士に使用してもらうことは、取調室導入の交渉においても不可欠であるとの当初の認識どおり、現場連携の実績を得ることができた。

問題発言のテキスト検出の難しさ、映像検出の現実的制約は概ね想定の範囲内であった。いずれも音響特徴・振動センサ・秘匿通知などの代替・補完アプローチで対応する方向性を確立できた。

装置としては背合わせの 2 モニタ、マイク、カメラ、振動センサ、深層学習用サーバを準備し、可搬装置

内での統合にはリアルタイム処理性能と小型化の両立が今後の課題として残る。

3 研究成果の総括と今後の展望

本研究調査により、取調室の言動解析と可視化に向けた以下の成果を得た。

(1) Whisper と LLM を組み合わせた問題発言検出の基盤手法と、音響特徴との統合設計を確立した。(2) 映像の限界を踏まえた振動センサ＋オートエンコーダによる異常行動検出手法を設計し、MediaPipe Face Mesh と Transformer による発話時顔部モーション識別（正解率 98.3%）の知見も得た。(3) 8 方式の秘匿通知評価実験により、取調室への警告提示に骨伝導イヤホンが最適であることを定量的に示した。(4) VIA ベースの統合可視化・半自動検証システムを構築した。第五に、Audio Scribe & Summarizer を弁護士 2 名の日常業務に定着させ、リーガルテックとしての現場連携基盤を築いた。

本研究の成果は、取調室に限定されない。病院での診察においては、難解な専門用語について患者側にだけ解説を付加する用途が考えられる。保育園や小中学校では、問題言動の検出と、問題がなかったことの証明の双方に資する。いずれの用途においても、セキュリティの確保と、撮影・記録される側にもメリットがあることが採用の条件となる。本研究で培った音声解析、センサ異常検知、秘匿通知、長時間データ可視化の各技術は、こうした幅広い対話モニタリング場面への展開が可能である。

今後は、模擬取調実験による総合検証、オフライン LLM 搭載機器の展開、国際論文誌・国内研究会での成果公表、取調室への本格導入に向けた弁護士会・関係機関との連携強化を進める。本研究で得られた手法は、病院での診察、保育現場、教育現場など、権力関係を伴う対話場面への応用も可能である。

助成期間内の投稿・発表は一部にとどまったが、秘匿通知評価の成果をまとめ、音声・センサ統合解析の論文化を今後も進める。取調以外の法律相談の内容をまとめる用途でも本技術はそのまま活用でき、Audio Scribe & Summarizer の継続利用はその証左である。違法・不当な取調の抑止と冤罪防止に向け、情報通信技術による安価で万人に対応可能なシステムの社会実装を引き続き目指す。

取調室というアクセスが困難な領域においても、音声解析、センサ異常検知、ヒューマンインタフェース（秘匿通知）、データ可視化の各技術を統合的に開発する機会を得た。本研究成果は、法曹界と情報通信工学の橋渡しとなるリーガルテックの新たな方向性を示すものであり、今後の政策議論や技術開発においても参照されることを期待する。

【参考文献】

[1] Dowell, N. M. and Berman, J. S.: Therapist Nonverbal Behavior and Perceptions of Empathy, Alliance, and Treatment Credibility, Journal of Psychotherapy Integration, Vol. 23, No. 2, pp. 158–165 (2013).

[2] Giebelhausen, M., Robinson, S. G., Sirianni, N. J. and Brady, M. K.: Touch Versus Tech: When Technology Functions as a Barrier or a Benefit to Service Encounters, Journal of Marketing, Vol. 78, No. 4, pp. 113–124 (2014).

[3] Logas, J., Belan, K., Lin, G., Gogate, A. and Starner, T.: Conversational Partner's Perception of Subtle Display Use for Monitoring Notifications, Augmented Humans International Conference 2021 (AHs '21), pp. 101–110 (2021).

[4] Hall, E. T.: The Hidden Dimension, Anchor Books (1990).

[5] Amlôt, R. and Walker, R.: Are somatosensory saccades voluntary or reflexive?, Experimental Brain Research, Vol. 168, pp. 557–565 (2006).

[6] Czerwinski, M., Cutrell, E. and Horvitz, E.: Instant Messaging: Effects of Relevance and Timing, People and Computers XIV: Proceedings of HCI, Vol. 2, pp. 71–76 (2000).

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
特定フレーズ発話時の顔の動きによるユーザ識別に対する顔部位ごとの有効性	画像の認識・理解シンポジウム (MIRU)	2026 年 7 月

Audio Scribe & Summarizer（会話文字起こし・要約・音声再生システム）	（山梨県弁護士会所属弁護士ほかへの実用展開）	2025 年 11 月～継続利用中