

「文字起こし」の効率化を考える

ENWライターチーム

Supported by EcoNetworks

2018年10月

問題意識

ライターの執筆業務では、取材音声を文字に書き起こす「文字起こし」作業が発生することがあります。「文字起こし」は、単純な作業である一方、音声を何度も聞き直さねばならないなど、非常に時間がかかります。

そこで今回、「**どうすれば文字起こしを効率化できるか**」を探るべく、以下の2つを実施しました。

- ①取材記録や文字起こしの実態に関するアンケート調査
- ②音声再生や自動文字化用のソフトウェア・アプリ比較

取材記録・文字起こしの実態: アンケート調査結果

実施期間: 2018年9月3日～9月12日

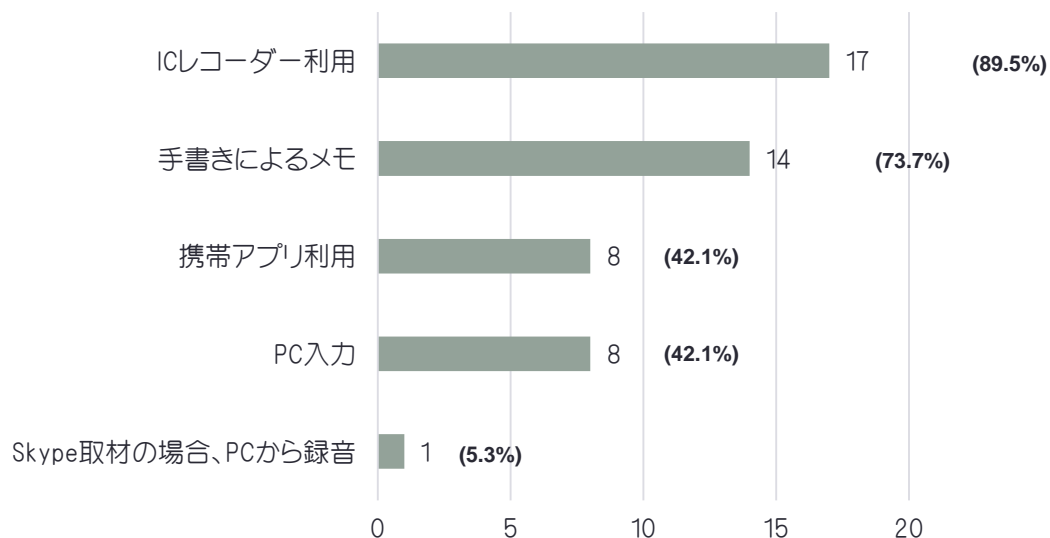
対象: ライター

回答数: 19

ICレコーダーの用途、クライアントのニーズ、文字起こしに使うソフトの有無などについて、日常的に文字起こしをしているライターにアンケートを行いました。

取材・ICレコーダーについて

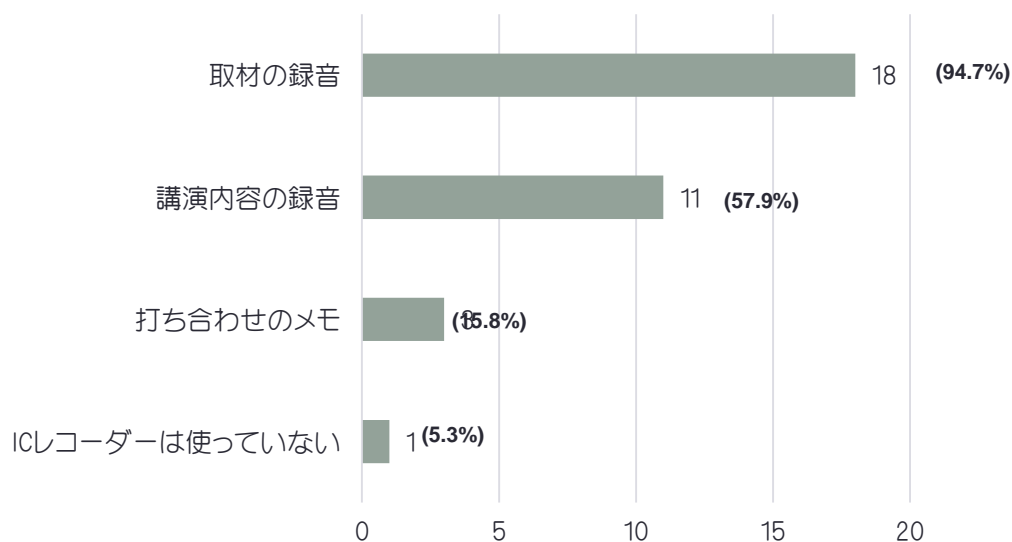
1) 普段、どのような方法で取材内容を記録・保存していますか？(複数回答可)



◆ICレコーダー派、手書きメモ派が多い(併用している場合あり)。

◆一方、PC入力や携帯アプリ利用者も多い。

2) 普段はどんなシチュエーションでICレコーダーを使っていますか？(複数選択可)

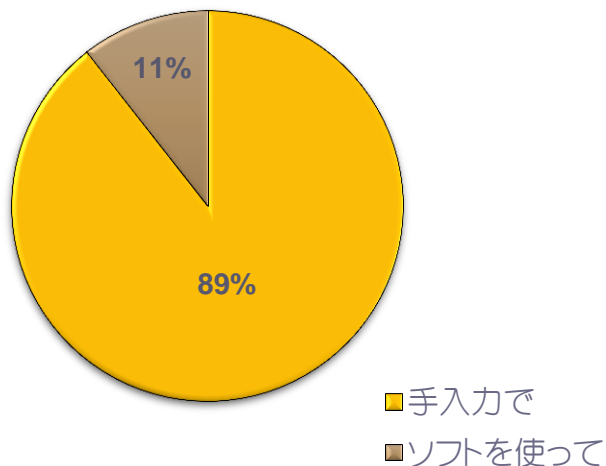


◆9割以上(19人中18人)が取材の録音に活用。

◆自身の音声メモとして使っている人は少数。

文字起こしについて

3) ICレコーダーの音源の文字起こしはどのようにやっていますか？下記からお選びください。

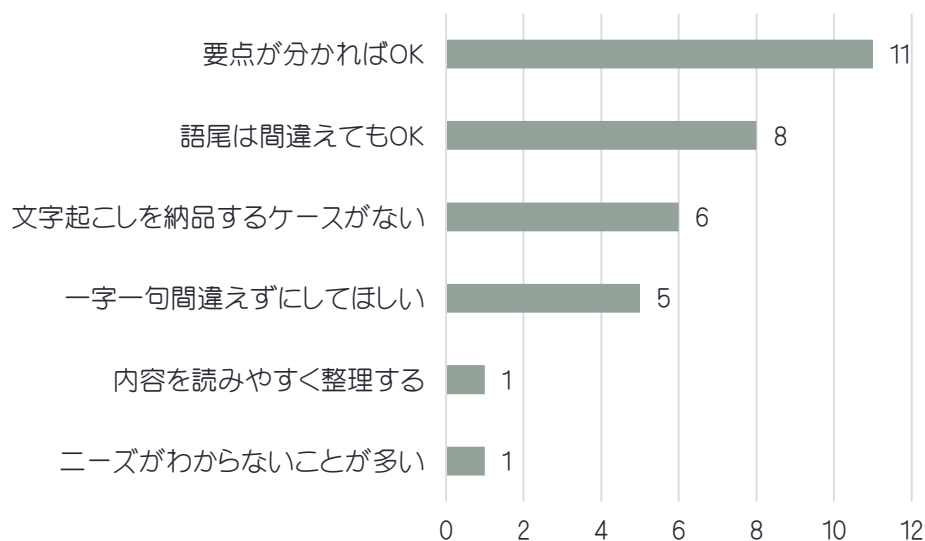


◆手入力で行っている人が19人中17人で、9割弱。

◆ソフトを使っていると回答した2人のうち1人は「音声再生サポートソフトを使って再生し書きとっている」と回答していた。これは「手入力」とほぼ同じ。

◆よって、ソフトで文字を起こしをしている人は19人中たった1人。

4) 文字起こしの際のクライアントのニーズを教えてください。

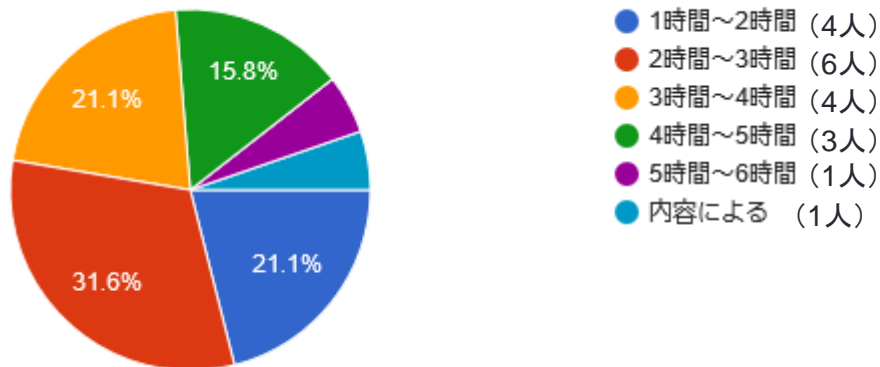


◆要点が分かればOK、語尾は間違えてもOKというケースが多い。

◆一方、一字一句間違えない文字起こしのニーズもある。

文字起こしについて

5) 所要時間について:1時間の音声データを文字起こしする場合、所要時間はどのくらいですか？



◆1時間の音声データにかかる時間は「2～3時間」という回答が最多(6人)。

◆2時間以内に終わる人は全体の2割程度の4人しかいない。その他は倍以上の時間がかかっている。音声データの文字起こしは時間がかかる作業だということが分かる。

■文字起こしの際、クライアントに明確にしてほしいことは？

- ・利用目的
- ・文字起こしだけなのか、起こしたあとまとめてほしいのか、その場合の文字数
- ・文字起こし原稿の用途。部内の記録用、役員に読ませる等がわかると、レベル感を合わせやすい

◆利用目的を明確にしてもらいたい、という声が多い。

文字起こしのコツ

■文字起こしのコツを教えてください。

《実際の文字起こしの方法》

- 0.7倍くらいの速度で再生しながら文字起こしをする
- 1/2のスピードで再生してとにかくそのスピードで文字を打ちまくる。慣れるとスピードをあげる、調子が悪いとスピードを下げる。これで話す時間から、起こせる文字数の時間は読めるようになります。(かなり集中力が必要ですが…)
- 基本的に止めず、「3秒戻る」とか「5秒戻る」をショートカットで繰り返しながら書き出す
- 1分ごとにとめる 5分ごとに巻きもどして間違いがないか確認する
- 1人1発言ごとに止めて書き出し、その場で重複などの内容は要約する／ワンフレーズごとに止めて、音声入力する。
- 手書きメモで全体を確認した上で、音声を聞きながら一文ごとに止めて、書き出す。
- PCのキーボードで音源を再生、送り、戻しを操作しながらタイピングする
- 2台目パソコンにダビングして、そこから再生し、15～30秒ごとにとめる(正確な文字記録が必要なとき)。一時停止のボタンに矢印を合わせておく。キーが大きいので、再生、停止がラクです。
- 結構丁寧にやる方だと思います。わからなかった語句をこの段階で調べたり、ノートと照らし合わせたり。ざーっと書き出し、その後聞き直す。効率化のために、わからないところを聞き直す作業は後回し。
- メモをベースに弱いところから書き起こす／不要なところは省く／一旦ざっと聞き流しながら書き出す／

《録音中(=取材・講演中)の配慮》

- 私の場合、インタビュー→文字起こし→原稿制作→原稿納品となるので、インタビュー中に文字起こしが必要な部分を明らかにしておきます。
- 自分で取材した内容で、おおまか流れはわかっているので、使うところだけを粗起こししているような感じです。人に頼むよりも、きっと時間はかかっているとは思いますが。
- 講演や取材などはある程度メモを取っておき、わからなかった部分をレコーダーを聞いて付加する。

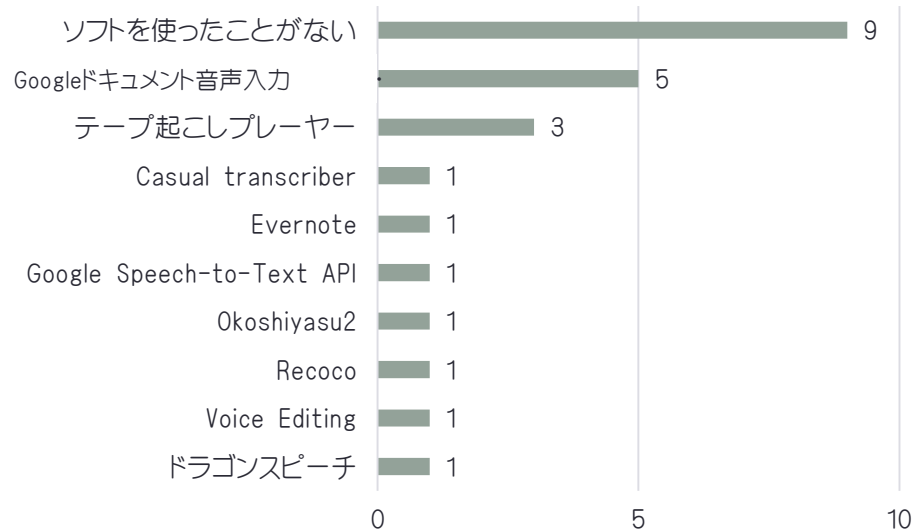
◆聞き流しながら書き起こし、あとで聞き返して間違いを修正していく、というスタイルの人が多い。

◆PCのキーボードの仕様をやりやすいように変えている人もいます。

◆録音時の配慮を教えてくださいました人もいます。

ソフトについて

6) 次の自動文字起こしや音声再生用ソフトを使ったことはありますか？
使ったことのあるソフトを選んでください。



◆全体(19人)のうち9人(47%)がソフトを使ったことがない。

◆Googleドキュメント音声入力は無料で使えるツールということもあり、5人が使った経験あり。次に多いのはテープ起こしプレイヤー。3人が使用経験あり。

文字起こしの改善策について

■ICレコーダーの文字起こし、こうできたら楽だなあ、というアイデアがあれば、教えてください。

《文字化機能について》

- 自由自在にゆっくりの速度にできるソフトを使えたらもっと楽。聞き取ったままそのまま書きだしてもらえたら、とても助かる
- 音声を再生しておくだけで自動的にテキスト化、ワード形式でダウンロードできるものがあるとうれしい。
- 自動文字起こしソフトで、大体が文字化されたうえで、聞きながら微修正できれば良い。
- 夜寝ている間に文字起こしが完了しているといい／録音した音源が、そのままテキストに変換される
- 取材から時間が経ってしまうことが多く、その人の呼吸や話し方を思い出すのに、もう一度音源を聞き直すことは今のところ省きたくない作業ですが、自動で正確に文字になってくれたらありがたいです。あと複数人の話し言葉を聞き分けてくれたら。
- 複数人が集まる座談会などの時、誰がどの発言か振り分けられたら、いいですねー。12人の時は起こしが大変でした。

《音声再生機能について》

- 一時停止、再生のボタンを軽く、大きく。(みなくても押せるくらいに)
- 1)再生スピードを変えられること、2)「数秒戻る」をショートカットでできることはソフト選びで必須でした。かなり楽です！
あと、ICレコーダーを置く場所で文字起こしにかかる時間が違うので重視したいです。

《発注元への要望》

- 雑音の少ないマイク音源であれば、自動文字起こしの精度がより高くなるので、なるべくマイクのようなものに音声は入れてほしい
(喫茶店などの場所でテーブルに置くタイプの安価なICレコーダーで撮った音声だと、自分の耳でも聞き取りにくい)
- クライアントによって、要求は全く異なりますのでなんともいえませんが、どちらかといえば、要点を抑えるだけでよい、という方が楽といえば楽です。難しい単語などの事前情報や会議時のメモがあるとありがたいです。

《その他》

- 誰かがやってくれたら楽だなあと思います

- ◆ソフトへの要望。音声を正確に文字起こしできるソフトがほしい、とうい要望が多い。
- ◆取材前・録音の際の要望。内容は録音環境や事前準備に関するリクエストとなっている。

まとめ: アンケート結果から見えてきたこと

- ICレコーダー音源の文字起こしは、手入力で行っている人が9割弱。ソフトを使っている2人のうち1人は、音声再生サポートソフトで再生して書きとっているため、手入力とほぼ同じ。つまり、文字を起こす部分でソフトを使っている人は、19人中たった1人。
- クライアントからのニーズは「語尾を間違えてもOK」「要点が分かればOK」が多い一方、「一字一句間違えてはならない」というニーズも25%ほどあった。
「クライアントにはニーズ、用途を明らかにしてほしい」という声が聞かれた。
- 1時間分の音声データの文字起こしが、2時間以内に終わる人は全体の2割程度(4人)しかいない。その他の8割程度は、2時間以上かかる。文字起こしには時間がかかる。
- 「使ったことのあるソフト」では、ソフトを使ったことがない人が9人。使ったことがある人も、再生サポートソフトを使っている。しかし「音声データを勝手に文字化してくれるソフトがあったらいいな」と思っている人は非常に多い。
- アンケート結果から浮かび上がるのは、多くのライターが「もっと楽にしたいと思いつつも、自力で長い時間をかけて、文字起こしをしている」実態である。

音声再生・自動文字化ソフトウェア／アプリの比較

アンケート調査では、文字起こしを手入力で行うライターが、9割に上りました。一方、ライターが自動文字化の精度アップを強く望んでいる実態も、浮かび上がっています。

AIなど技術革新が進む中、今後は音声の再生や自動文字化をサポートするソフトウェアを活用することも考えられます。



そこで、いくつかの音声再生・自動文字化ソフトウェアやアプリの比較を行いました(※)。それぞれの機能や特徴、使用した所感は、次ページ以降を参照してください。

※条件: ソフト試用にあたっては、机の上に置いたICレコーダーの内蔵マイクで録音した取材音声データを使用。なお、試用できていないソフト・アプリも、比較表に一部含まれる。

音声再生・自動文字化ソフトウェア／アプリの比較

	価格	利用方法	音声再生機能	自動文字化機能	コメント
Writer.app (無料版)	無料	外部の音声出入力ソフト(無料)をインストールした上で、オンラインからWriter.appを利用	<ul style="list-style-type: none">•音声入力(Chromeのみ)•音声再生(再生速度・音量などの調整)•テキストエディタ	<ul style="list-style-type: none">•録音データやマイク入力音声を、自動文字化•ボタン操作で、自動改行、箇条書きやコメント形式の表示などが可能•機械学習機能	<p>【使用した所感】</p> <ul style="list-style-type: none">•取材録音データは、音声認識はされるが、大半が誤認識・誤変換。文章化の精度は、低い。•静かな環境で、マイクに向かってゆっくり明瞭に話すと、きちんと音声認識・文字化される。•外部ソフトのインストールにより、PC環境によっては動作が遅くなることもあった。

音声再生・自動文字化ソフトウェア／アプリの比較

	価格	利用方法	音声再生機能	自動文字化機能	コメント
Writer.app (有料版)	月額 500円 +従量 課金 1分8円	インストール 不要、オンラ イン利用	<ul style="list-style-type: none">•音声入力•音声再生 (再生速度・音量 などの調整)•テキストエディ タ	<ul style="list-style-type: none">•Google Speech-to- Textによる音 声データの自 動文字化•画像の文字起 こし•音声／執筆 データの保存・ 管理•機械学習機能	<p>【未使用】</p> <ul style="list-style-type: none">•インストールなし で、オンラインで 利用・管理できる ため便利•静かな環境で、マ イクに向かって ゆっくり明瞭に話 し、近くから録音さ れたデータが必要

音声再生・自動文字化ソフトウェア／アプリの比較

	価格	利用方法	音声再生機能	自動文字化機能	コメント
VoXT SELF (ボクスト・セルフ)	ソフトは無料、利用は従量課金 1分30円	無料ソフトウェアをインストール	<ul style="list-style-type: none"> •音声入力 •音声再生 	<ul style="list-style-type: none"> •録音データやマイク入力音声、自動文字化 •文字化された文章結果を、該当音声を再生しながら、確認・修正ができる •音声／執筆データの保存・管理 •各作業完了ごとに通知メール 	<p>【使用した所感】</p> <ul style="list-style-type: none"> •取材録音データは、音声認識・自動文字化は、40-70%程度の精度(音声品質、専門用語の量による)。 •誤認識・誤変換もあるため、別途自分で音声をすべて確認・修正する作業あり。 •1時間の音声を、15分程度で文字化 •作業完了通知メール機能やワード保存機能が便利

音声再生・自動文字化ソフトウェア／アプリの比較

	価格	利用方法	音声再生機能	自動文字化機能	コメント
Speechnotes	無料	オンライン利用(Chrome)	<ul style="list-style-type: none">・音声入力・音声再生・テキストエディタ	<ul style="list-style-type: none">・録音データやマイク入力音声を、自動文字化・音声／執筆データの保存	<p>【使用した所感】</p> <ul style="list-style-type: none">・取材録音データの場合は、半分程度の音声しか認識できず、文章として成り立たない。誤変換が多い。・マイクに向かってゆっくり明瞭に話すと、きちんと音声認識・文字化される。

音声再生・自動文字化ソフトウェア／アプリの比較

	価格	利用方法	音声再生機能	自動文字化機能	コメント
Google Speech-to-Text API	60分まで無料。以降、1分0.024USD (約2.4円)	<ul style="list-style-type: none">・オンライン利用 (Chrome)・コード入力などプログラミングが必要	<ul style="list-style-type: none">・音声入力・音声再生・テキストエディタ	<ul style="list-style-type: none">・録音データやマイク入力音声を、自動文字化・120言語対応・音声／執筆データの保存・管理・機械学習機能	<p>【未使用】</p> <ul style="list-style-type: none">・Google Cloudプラットフォーム上で管理できる。・使用にあたっては、コード入力によるプログラミングが求められるため、初心者には難しい場合がある。・mp3/mp4の音声データに対応していない (FLAC/WAVなどに変換する必要あり)

音声再生・自動文字化ソフトウェア／アプリの比較

	価格	利用方法	音声再生機能	自動文字化機能	コメント
Googleドキュメント音声機能	無料	Googleドキュメントを利用	なし	・マイク入力音声を、自動文字化	【使用した所感】 ・取材録音データを再生しても、音声認識・文字化はされなかった。 ・PCに向かってゆっくり明瞭に話すと、ほぼ正しく音声認識・文字化される。

音声再生・自動文字化ソフトウェア／アプリの比較

	価格	利用方法	音声再生機能	自動文字化機能	コメント
テープ起こしプレイヤー	無料	無料ソフトウェアをインストール	<ul style="list-style-type: none">•音声再生•再生速度調整(0.5-2倍速)、音量調整・加工などの機能•ショートカットキーによる再生停止、早送り、巻き戻し、ステレオ音声のモノラル化	なし	<p>【使用した所感】</p> <ul style="list-style-type: none">•2倍速にすると聞きながら書き続けられる点、ショートカットキーを設定できる点が便利。•自動文字化機能はないので、それなりの労力と時間は必要(5分の音源／2倍速の文字起こしに、約25分かかった)。

ソフト・アプリ比較から見えてきたこと

- 今回比較したソフトのうち、自動文字化の精度が比較的良かったのは、「VoXT SELF」であった(40-70%程度。ただし、録音データの音質、専門用語の量によってかなり差が出る)。
誤変換・誤認識も多いため、改めて自分で音声をすべて聞きながら、微修正する作業が発生する。
- 今回比較したソフト・アプリの大半で、ICレコーダーで録音された取材音声は、現時点で自動文字化の精度は低く、文章として成り立たないことが多い。
(理由:取材・講演録音データは、話すスピードがより速く、複数人が話すこともある上、一定の距離からICレコーダーなどで録音されるため。)
- ただし、これらのソフトでも、静かな環境でゆっくり明瞭に話し、マイクに直接入力された音声であれば、より良い精度で認識・自動文字化しうる。取材録音時には、ICレコーダーに外部マイクを接続したり、専用マイクシステムを使用したりすることで、自動文字化の精度を上げられる可能性はある。
- その他、音声再生ソフトで再生速度を落とし、ショートカットキーによる巻き戻し機能などを使うことで、文字起こしの労力・時間を減らすことは可能。
- 取材・講演録音データの文字起こしについては、録音環境に十分注意して改善しながら、①「VoXT」などの比較的精度の良いソフトで、自動文字化機能を活用する、②「テープ起こしプレイヤー」などの無料音声再生ソフトで、再生速度を落としながら、自力で効率的に文字化することが有効と考えられる。今後、自動文字化技術のアップデートに期待。

参考: 自動文字化ソフト「VoXT SELF」のトライアル結果

【事例①】 会議室でのインタビュー音声

録音環境: 静かな会議室内で、テーブルを挟んでの取材。発言者はメイン1名、その他3名。

ICレコーダーは、取材相手から1mの場所に置く。

音声品質: 若干早口ではあるが、比較的明瞭な音声

自動文字化の所要時間: 1分のデータ、音声の文字変換5分

自動文字化結果: 音声認識・変換ともに7割くらいの精度。音声のクリアさに加え、専門用語が少なめだったため、精度が比較的良かった可能性あり。

【事例②】 会議室でのインタビュー音声

録音環境: 静かな会議室内で、テーブル2つを挟んでの取材。発言者はメイン1名、その他3名。

音声品質: 比較的明瞭だが、ICレコーダーから遠い人物の声は聴きとりづらい。

自動文字化の所要時間: 1時間のデータで、データ保存10分、音声の文字変換15分

自動文字化結果: 音声認識・変換ともに4割くらいの精度。専門用語が多かったことが、精度低下の要因の一つか。

【事例③】 テレビ会議システムを通じてのインタビュー音声

録音環境: 会議室内でテレビ会議システムを通じて取材。発言者はメイン1名、その他2名。

ICレコーダーは、テレビ画面から1-2mの場所に置く。

音声品質: 音声がくぐもって若干響いている状態 (テレビ会議音声でやマイクの位置が要因か)

自動文字化の所要時間: 1時間20分のデータで、データ保存25分、音声の文字変換15分

自動文字化結果: ほとんどが誤認識・誤変換で、文字起こしとしては使えなかった。

(→テープ起こしプレーヤーで、0.7倍の速度で再生しながら文字起こし。所要時間4時間弱)

まとめ: 今後の文字起こしの効率化のヒント

- **録音前:**

録音環境に気を配り、録音の精度を上げる →精度によって、自動文字化機能の活用も可。
(ICレコーダーに外部マイクを接続しての録音、マイクシステムの活用、ICレコーダーを置く距離、話すスピード、静かな場所等)。

- **文字起こし作業前:**

あらかじめクライアントに、文字起こしの利用目的を確認し、求められる精度を把握した上で作業する。

- **文字起こし作業:**

① 比較的精度がいい自動文字化ソフト (「VoXT SELF」等) を活用する。ただし、誤認識・誤変換も多いため、自分自身でも再度音声を聞きながら修正する必要がある。ソフト利用料に加えて、こうした追加作業に費やす労働時間を踏まえて、ソフトを活用するか、自力で文字起こしをするかを検討する。

② もしくは、音声再生ソフトを活用する (「テープ起こしプレイヤー(無料)」等)。

(音声速度を下げ、聞きながら書き起こす/ショートカットキーの設定により、巻き戻しなどの動作を簡単にするとよい)

- **通常業務として:**

今後、自動文字化技術の進歩が期待されるので、各種ソフトのアップデートをチェック。

- **その他:**

文字起こしを請け負う会社やフリーランサーも多くあるため、予算に応じて外注も検討。