



議事録作成支援サービス

COTOHA Meeting Assist

# 音声テキスト化の精度について

2021年 2.4版  
NTTコミュニケーションズ株式会社

はじめに

日本語は、文法も複雑で同音異義語が多く、音声認識技術としては難しい言語です。しかし、ちょっとした收音のコツに気を付けることで、音声認識精度を向上させることが可能です。

本ドキュメントでは、COTOHA Meeting Assistでの音声認識がうまくいかないときに、ご利用環境をチェックしていただくポイントを解説します。

## ■ 音声認識精度を向上させる4つのチェックポイント

**Point 1**    マイクの選定 . . . . . 指向性のあるマイクかどうか

**Point 2**    入力音量の調節 . . . . . 音量が適切かどうか

**Point 3**    マイクと発話者の距離 . . . 距離が適切かどうか

**Point 4**    話し方 . . . . . 語尾まではっきり発音しているかどうか

**次ページから詳細をご説明します。**

## サービスの基本的な利用方法

会議を行う際は参加者毎にそれぞれのパソコンまたはスマートフォンからお客様専用サーバーに**個別のユーザーID**でログインして会議に参加します。



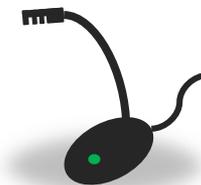
- 発言内容は、**音声が入力された端末のログインユーザーの発言**として発言毎に記録されます。
- ログインする端末と外部接続マイクは**発言者毎に必要です**。

### 準備が必要なもの 参加者全員

クライアント端末  
(パソコン または スマートフォン)



外部接続マイク  
(指向性マイクを推奨)



## 音声テキスト化の精度について

**!** COTOHA Meeting Assist は議事メモ作成を支援する補助ツールです

音声データから生成されるテキストは、**実際に発言された内容であることを保証するものではありません**。なお、音声データ変換精度は処理を行う音声データの品質により影響を受けるため**ご利用環境やご利用方法によって変動します**。

### 正しくテキスト化されない例



**▶ 音声テキスト化の精度は次項からご案内するご利用方法によって改善が期待できます。**

音声テキスト化の精度を向上させるポイント

生成されるテキストの精度は入力される音声品質に大きく影響されます。  
音声品質が良いほど認識精度が高くなります。

音声品質は入力時に決まる



マイク



クライアント端末



COTOHA  
Meeting Assist

提供サーバー

音声入力

音声のテキスト化

▶ テキスト化の精度を高めるには  
入力音声の品質を高めることが重要です。

入力音声の品質を向上させるポイント

## Point 1 マイクの選定

マイクには特性があり集音範囲が異なります。一つのマイクに周囲の雑音や他人の発言が同時に拾われることで発言内容だけを正しく抽出できずテキスト化の精度が低下します。

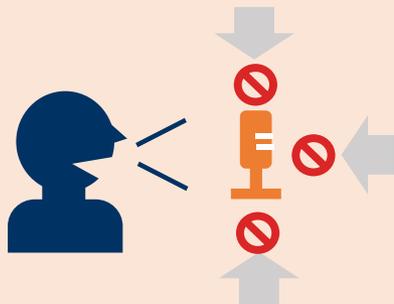
### 単一指向性マイク

特定の方向への感度が高い

○ 他人の発言や雑音を拾いにくい  
ため適切にテキスト化される

例

(発話内容) 海外出張に行くことになりました  
(認識結果) 海外出張に行くことになりました



### 全指向性マイク

すべての方位から同等の感度

✕ 他人の発言や雑音を拾い発言内容が正しく  
テキスト化されないことがある

例

(発話内容) 海外出張に行くことになりました  
(認識結果) 貝が移出コンテナ



▶ 単一指向性等のマイクを用いることで集音範囲を限定し、音声品質を向上させることができます。

入力音声の品質を向上させるポイント

## Point 1 マイクの選定

基本的には、マイクの機種・メーカーは問いません。  
例として、弊社で利用実績のあるものをご紹介します。

### 推奨スタンドマイク

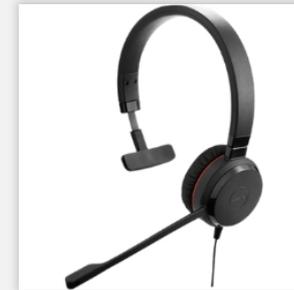


メーカー : サンワサプライ  
品番 : MM-MCUSB25  
接続IF : USB

<参考情報>

NTT-X Storeでの販売価格  
2,519円（税込）

### 推奨ヘッドセット



メーカー : Jabra  
品番 : Evolve 30 II MS Mono  
接続IF : USB

<参考情報>

NTT-X Storeでの販売価格  
5,450円（税込）

入力音声の品質を向上させるポイント

## Point 2 入力音量の調節

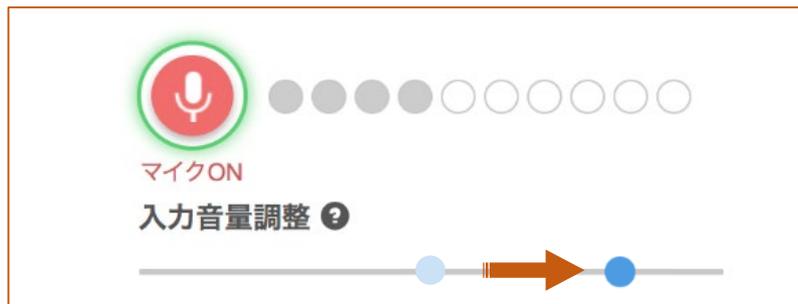
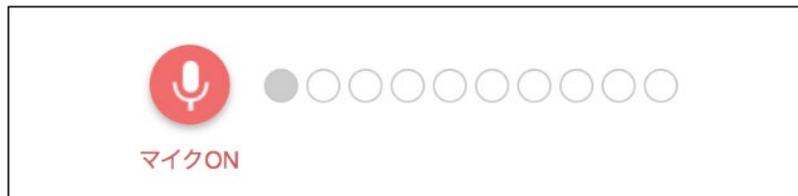
マイクから入力される音声の音量によりテキスト化の精度が異なります。  
小さすぎると音声を認識できず、大きすぎても周囲の音を拾ってしまい正しく認識できません。

コントロールパネルのマイクボタンの周囲が **緑色** に光る状態が最適です。

「**マイクON**」の状態音声入力した際に以下のような状態の場合は、入力音量の調整をお願いします。

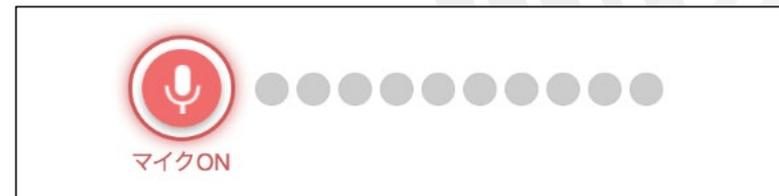
### A) マイクボタンの周囲の色が変わらない

音量が不足しています。入力音量調整のスライダーを右に移動させてください。



### B) マイクボタンの周囲が赤色に光る場合

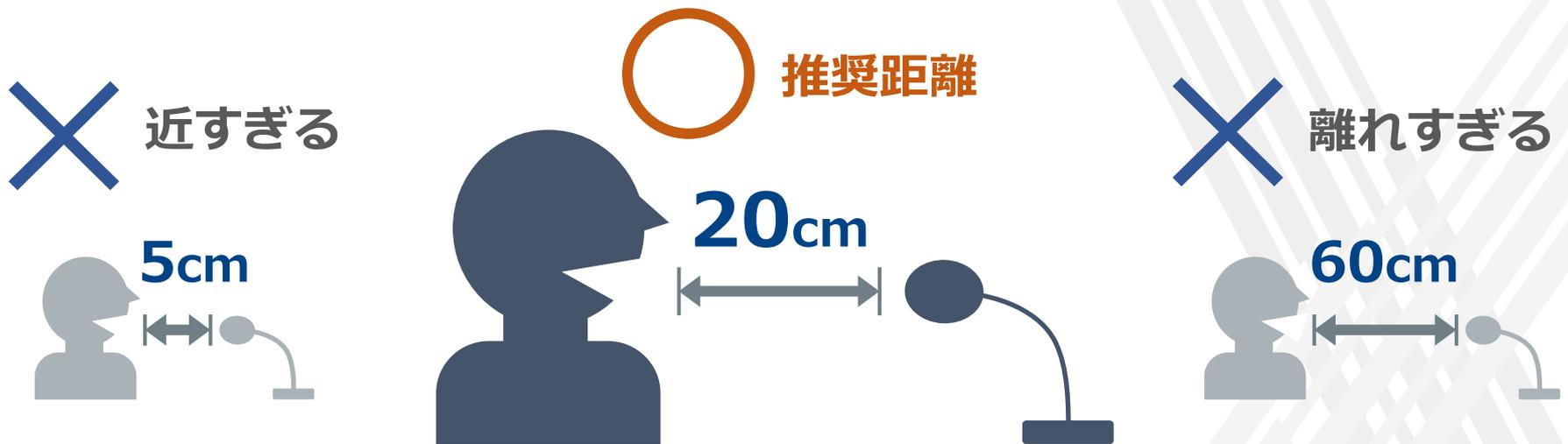
音量が大きすぎます。入力音量調整のスライダーを左に移動させてください。



入力音声の品質を向上させるポイント

### Point 3 マイクと発話者の距離

マイクと発話者の距離によって入力音声の品質に差がでます。  
近すぎても音声割れが起きてしまい正しくテキスト化できず、離れすぎても音声をマイクが拾えずテキスト化できません。  
例えば、Point1でご紹介したスタンドマイクの場合は、20cm程度離れた距離が最適です。



▶ マイクと発話者の距離を推奨距離にすることで  
テキスト化の精度が向上します。

入力音声の品質を向上させるポイント

## Point 4 話し方

話し方により認識精度に差が生じます。

特に無声音（「き / く / し / す / ち / つ / ひ / ふ / ぴ / ぷ」など）含む言葉は、テキスト化の際に誤認識されやすいです。



- 語尾まで丁寧に話す
- はっきり発音する



- 語尾をはっきり発音しない
- ぼそぼそ話す
- 極端に小さい声で話す

正しくテキスト化できない例

- 文章が欠落する  
発話内容) はいそうです  
認識結果) はいそうで
- 音声 that 認識されない  
発話内容) 新大阪と東京間を  
認識結果) (空白)

はっきりとした発音で語尾まで丁寧に話すことで  
テキスト化の精度が向上します。

