



# 研究を『本気で』製品にする ～とある音声認識研究者の場合～

2024年3月5日

NTT人間情報研究所

浅見 太一

- これまで、音声認識研究者として、また、サービス開発部門のマネージャとして、自分で研究をして、成果をプロダクト化する機会を何度か得ることができました
- 研究成果をプロダクト化する過程で苦労したことは多々ありますが、特に自分の中で印象に残っている、糧になっていると感じているものを今日はお話しさせていただきます
- 個人的な経験談ではありますが、研究の位置づけに悩んでいる方、研究のその先がイメージできずモヤモヤしている方に、何か一つでも視点を提供できれば幸いです

# 自己紹介



## 浅見 太一 (あさみ たいち)



### ◆ 所属

- 人間情報研究所 思考処理研究プロジェクト 主幹研究員

### ◆ プライベート

- 趣味： ラーメン屋巡り、音楽鑑賞(HR/HM)
- 住居： 東京 (武蔵小山)

### ◆ 経歴

- 2003.4： 東京工業大学 古井・篠田研に配属
  - 話者認識の研究、**音声**処理との出会い
- 2006.4： NTT研究所に入社 (研究員)
  - 音声**認識の研究開発、国際会議 (ICASSP, INTERSPEECH) 発表8回
  - 衆議院議会録システム、コールセンタ通話分析システム
- 2017.8： NTTドコモに転籍 (担当課長)
  - 音声**サービスの基盤開発、サービス運用
  - 151音声認識IVR、my daiz、AI電話
- 2020.7： NTT研究所に戻る (グループリーダー)
  - 研究グループのマネジメント



AI電話サービス

# 今日お話しするプロダクト



## ■ NTT-TX : ForeSight Voice Mining (2014年)

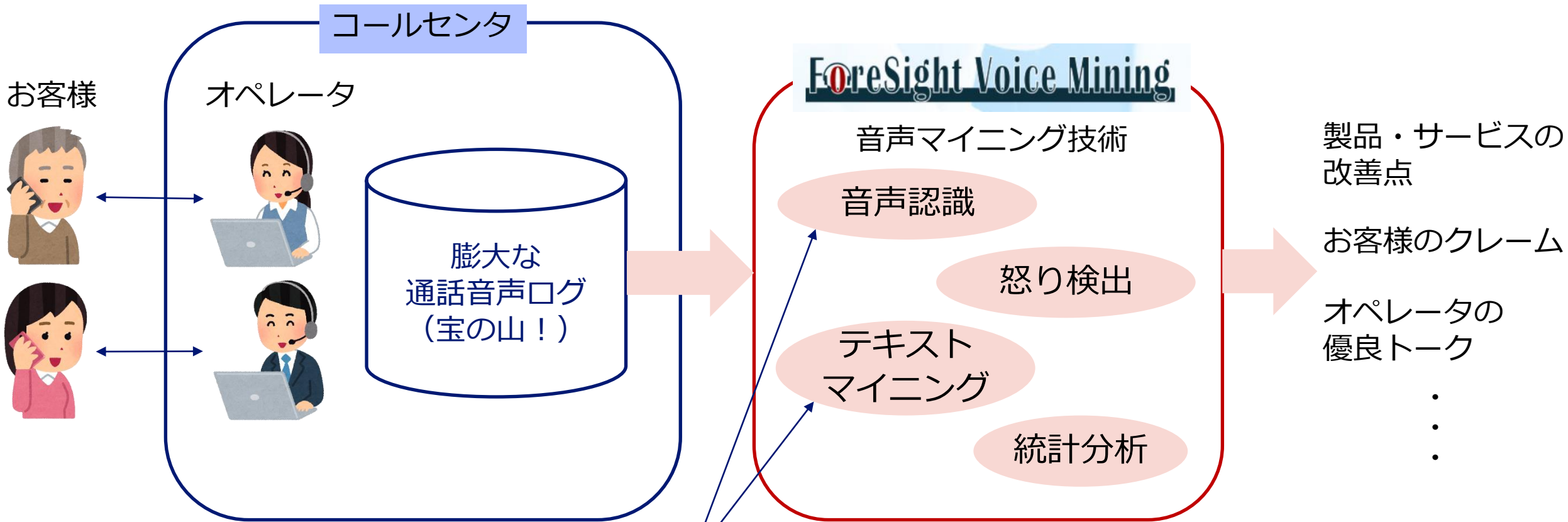
ForeSight Voice Mining

- コールセンタに蓄積される「お客様の声」を音声認識で全てテキスト化
- 定量的・客観的に分析して、サービス改善に繋げるソリューション

## ■ NTTドコモ : 151音声認識IVR (2017年)

- 8500万人超のドコモユーザからの電話を受け付ける総合受付センタ151
- 電話をかけて「6」をプッシュすると、音声認識でご用件を伺い、自動で回答したり、適切なオペレータに割り振る

# ForeSight Voice Mining (FSVM)



私が研究していたところ

- ・ 音声認識の精度向上
- ・ 認識誤りに頑健な分析方法

「私の（我々の）研究はこんなにも無力なのか」

- 音声マイニングの研究を2007年に開始
  - 試作システムを開発し、社内コールセンタで実験
  - 音声認識
    - › オペレータ認識率：66%、お客様認識率：47%
    - › 現実の会話音声にはまったく歯が立たなかった・・・
  - 怒り検出
    - › お客様が怒鳴っている場合は検出できる
    - › 実際は怒鳴っているお客様（Hot Anger）は非常に少ない
    - › 怒鳴らず、淡々と怒りをぶるけるお客様（Cold Anger）が検出できない・・・

価値を信じられるなら、やめない

## ■ FSVMの製品リリースは2014年

- 音声認識
  - › 毎年1~2%ずつ、地道に精度を高めた
    - » 推論アルゴリズムや学習データを少しずつ改善
  - › 2007年には想像できなかった技術の進歩もあった（深層学習、教師なし適応）
- 怒り検出
  - › お客様の音声だけからCold Angerを検出することは難しい
    - ⇒ 怒りをぶつけられているオペレータの音声にも特徴が現れるのでは
  - › オペレータの会話の特徴（「間」が長くなる等）を利用した新方式を開発
- コールセンタの現場の方から「まだやってたんだ」と言われながら研究

## 技術の使い方を誰も知らない

- コールセンターの通話音声ログを分析するシステムを開発しました！  
⇒どう分析したら業務改善できるの???
- (当然ですが) 「通話を分析して業務改善した経験がある人」はいません
  - › 当時はデータアナリストはメジャーではなかった・・・
- 技術の価値を理解 (実感) してもらえず、導入にGoサインが出ない
  - › 導入にはコストがかかるので、これも当然の判断



## 「最初の一例」を作るのは開発者の役目

- 共同実験先の社内コールセンタにシステムを設置し、自分で分析
  - 「某サービスの改善により〇千万円のコスト削減に繋がる」を突き止め、最初の導入先となった
  - もちろんこれだけが決め手ではありません
  - 「新技術の導入に積極的な人たち」と仲間になることもとても重要
- 新しい楽器の最初の演奏者は開発者（難しい／クセのある楽器なら）

# 151音声認識IVR

お客様



新料金プランについては4→6  
キャンペーンについては4→5  
紛失・発見は1  
設定方法は2  
故障・電波状況は3  
NMTは4  
...

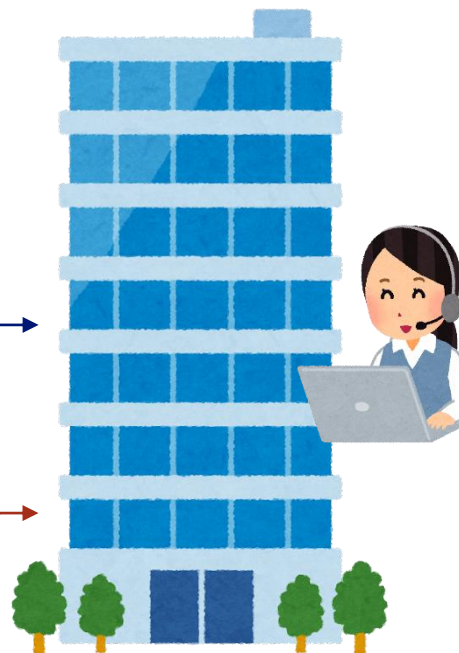
IVR

音声認識でご用件をお伺いしますので6を...

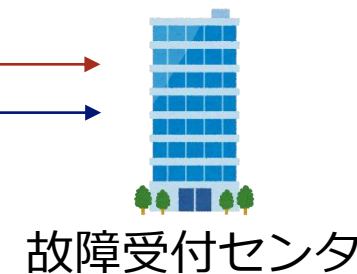
音声認識

意図解釈  
対話

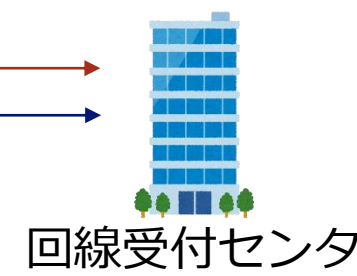
総合受付センター



専門センター



故障受付センター



回線受付センター



サービス問合せセンター

私が携わったところ

- ・音声認識の精度向上 (研究所)
- ・システム全体の開発 (ドコモ)

# 151音声認識IVR: 苦勞したこと①



どんな入力にも対応する必要がある

## ■ お客様は千差万別

- 老若男女
- 用件の話し方
  - › 「料金プランについて教えてください」
  - › 「新料金プラン」
  - › 「あー、えーっと、電話料金のことが、聞きたいんだけどー」
- 使う電話・回線、場所
  - › 異なる圧縮方式
  - › 静かな部屋、駅の雑踏、・・・

## 地道に、全て対応する

- 音声認識・対話の学習データをとにかく拡充
  - トライアル（お試し公開）を実施して、実際のユーザの発話例を収集
  - 話し方のバリエーションを増やして数万パターンの入力文例を作成
  - 作成した文例を様々な人に喋ってもらい、音声を集
  - 様々な圧縮方式や雑音を音声に適用
  - 作ったデータでモデルを学習
- 入力として有り得るデータをひたすら集める・作る

## 変化に対応し続ける

- リリースした後も、入力は変わり続ける
  - 毎月登場する新サービスの名前
  - それまでなかった言い回し（「ギガが減る」）
  - 新機種の音声圧縮方式
  - 使われなくなる言葉もある
- 放っておくと、IVRが正しく振り分けられなくなってしまう

## チームで対応する

- 社内の様々な部署から情報収集して継続的に対応
  - サービス企画部門 →翌月の新サービスの名称・略称
  - システム運用部門 →IVRの振り分けログに異常がないか
  - 製品開発部門 →新機種の声処理方式
  - 毎月モデルをアップデートして対応
- どんな情報を集めれば良いか、対応が必要かどうかを目利きできるのは音声認識の中身を知っている人だけ

# 「F1の研究」と「乗用車の研究」



どちらも大切

自分が今どちらをやっているか自覚することも大切

## ■ F1

- 決まったレギュレーションの中でstate-of-the-artを目指す
- コストは度外視
- プロが扱えれば良い

## ■ 乗用車

- どんな道でも安定して走る、アクシデントに対応する
- コストは非常に重要（安く作れるようにする）
- 素人が扱えるようにする

- なぜプロダクト化の苦労・努力をするのか

自分のアイデア・研究で人の役に立ちたい、喜んでもらいたい

- 何か一つでも心に残る話ができているだけで幸いです