

# 人とAIがともに音楽を作り出す未来に 向けた歌声合成の取り組み

ヤマハ株式会社 研究開発統括部  
才野 慶二郎



# 本日の概要

## 自己紹介

才野 慶二郎 (41)

- ヤマハ株式会社 研究開発統括部
- 歌声合成技術開発など



## 今日の話題

- 『研究がプロダクトになるまで』
- 「生成AI」の使い方に関するヤマハ株式会社の事例紹介
- 特に、今日は歌声合成に関する部分にフォーカスしてご紹介

# Outline

- 歌声合成技術とは何で、ヤマハは何を目指しているのか
- 具体的な、未来を見据えた歌声合成研究の取り組み
  - 具体例1：VOCALOID  $\beta$ -STUDIO
  - 具体例2：TransVox（なりきりマイク）
- むすび

# Outline

- 歌声合成技術とは何で、ヤマハは何を目指しているのか
- 具体的な、未来を見据えた歌声合成研究の取り組み
  - 具体例1：VOCALOID  $\beta$ -STUDIO
  - 具体例2：TransVox（なりきりマイク）
- むすび

コーポレートスローガン

感動を・ともに・創る

## 楽器

- アコースティック楽器
- **デジタル楽器**
- etc



「音源技術」の  
挑戦の歴史

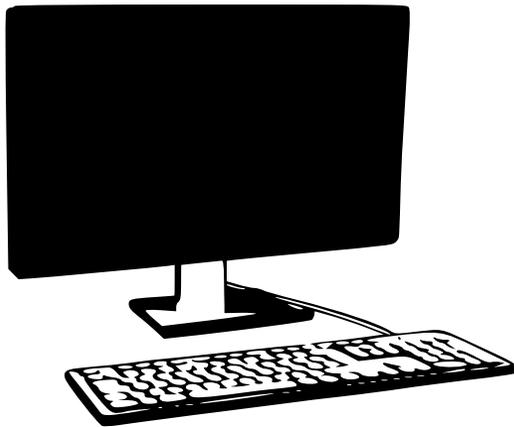
## 音響

ほか

# DTM (Desktop Music)

DTM = 生演奏ではなくPCの中で行われる音楽制作

電子楽器の音源技術により、**演奏**だけでなく**制作**が大きく発展



「最後に  
残された聖域」

「伴奏は作れるがボーカルは作れない」時代が長く続いた

# ヤマハの歌声合成にかける思い

人間の歌声という非常に魅力的な音楽表現を  
どうにかしてデジタルで実現したい

この想いは技術開発開始当初から現在に至るまで変わっていない  
時代ごとに、そのとき使える技術を駆使した研究開発を継続

# Outline

- 歌声合成技術とは何で、ヤマハは何を目指しているのか
- 具体的な、未来を見据えた歌声合成研究の取り組み
  - 具体例1：VOCALOID  $\beta$ -STUDIO
  - 具体例2：TransVox（なりきりマイク）
- さいごに

# 具体例1：VOCALOID $\beta$ -STUDIO

従来のDTM歌声合成のあたり前を改めて問う歌声合成の提案



# VOCALOID $\beta$ -STUDIO

What is this?

VOCALOID製品開発（現在）と並行して行う研究開発（未来）の一環

製品としての  
VOCALOID  
実装開発  
（現在）

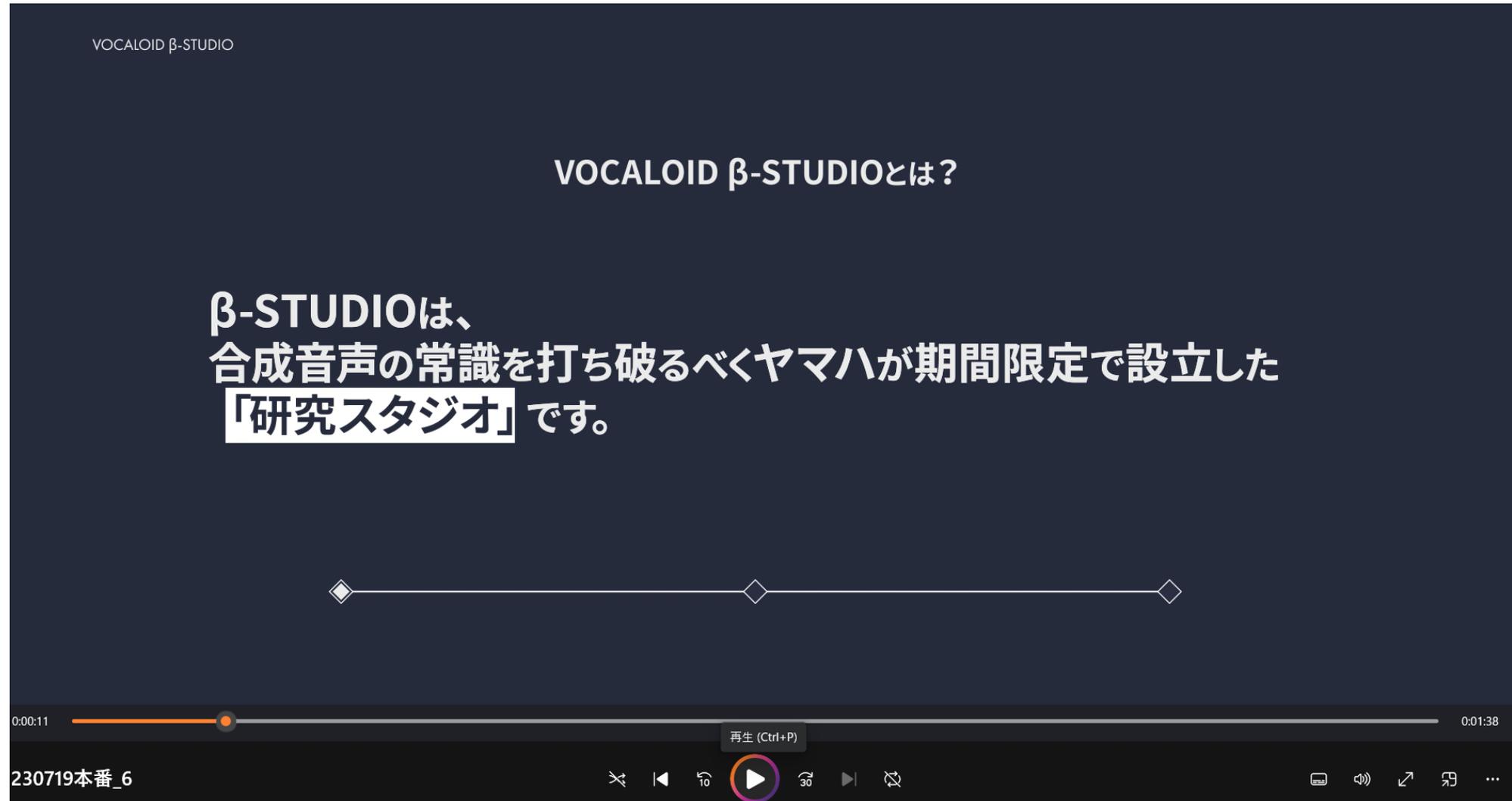


技術としての  
VOCALOID  
研究開発  
（未来）



# VOCALOID $\beta$ -STUDIO

## コンセプトムービー



# VOCALOID $\beta$ -STUDIO

## Why?

- AI時代のDTM用歌声合成ソフトの在り方を模索し、オープンに問うための『研究スタジオ』
- 研究段階の試作ソフトを数量限定で無償配布（不定期抽選）
  - 想定ユーザーは音楽表現の意思・意図を持ったクリエイター
  - 参加者様への期待は機能要望というよりは、とにかく音楽を生み出すこと
- 2023年8月22日に始動、2024年3月31日までの期間限定プロジェクト

# VX-β の特徴



AIによってもたらされる価値が  
最大化されるような試作

## 【合成エンジン】

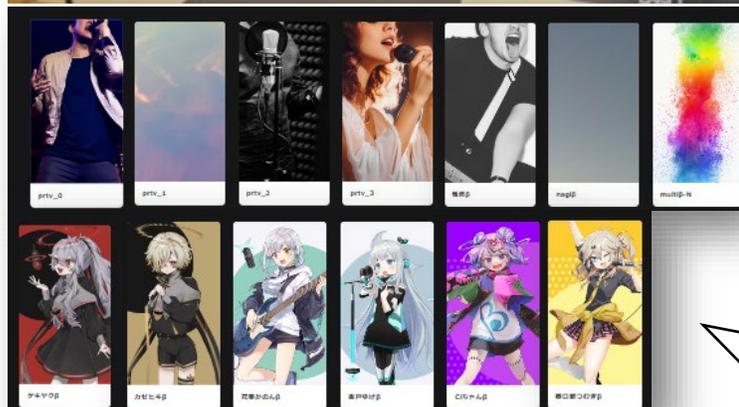
製品VOCALOIDと比べて**元歌唱者らしさ**をより強く表現するようなチューニング

## 【UI/UX】

一般的な楽器音シンセのような**プラグイン**として動作（世の中のあらゆる歌声合成ソフトの常識&制約「専用エディタ＝ピアノロールが必須」を疑うアプローチ）

## 【ボイスバンク】

多くのユースケースをカバーするように**幅広く・多数**搭載（ナチュラル寄り⇔キャラ寄り、男声⇔女声、日⇔英、絵あり⇔絵なしなど）



VX-β 同梱  
ボイスバンク

# 研究と製品のはざま

## 製品開発

- 既に形になって走っている中で大胆な変更は大変

## 研究

- 自由度は高いが、製品としての落とし込みは大変
- 「技術的強み」「ユーザーニーズ」のにらみ合い状態（?）

## VOCALOID $\beta$ -STUDIO

- 既存製品にあえて縛られない形で研究的・挑戦的な試作を公開
- 「おいしいところ」はユーザーに発見してもらおうという姿勢
- 今年度いっぱい（今月末まで!）の活動の知見を踏まえて今後の製品開発につなげていく

# 具体例2：TransVox™

音声入力による歌声合成（≒声質変換）



# TransVox

What is this?

音声入力による歌声合成（≒声質変換）

特長

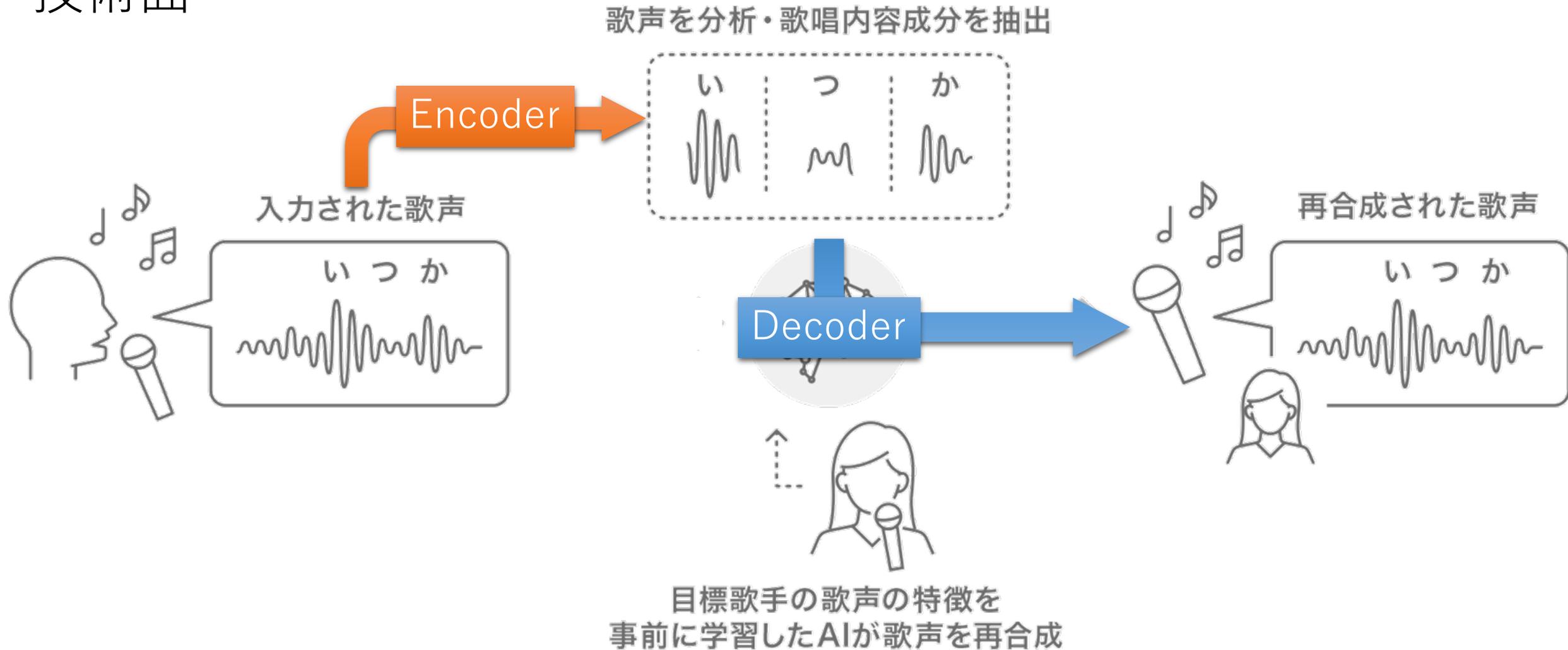
「カラオケルームで**楽しく**使える！」

- 低遅延（理論値で48ms）
- 雑音 = BGMまわりこみなどへの耐性

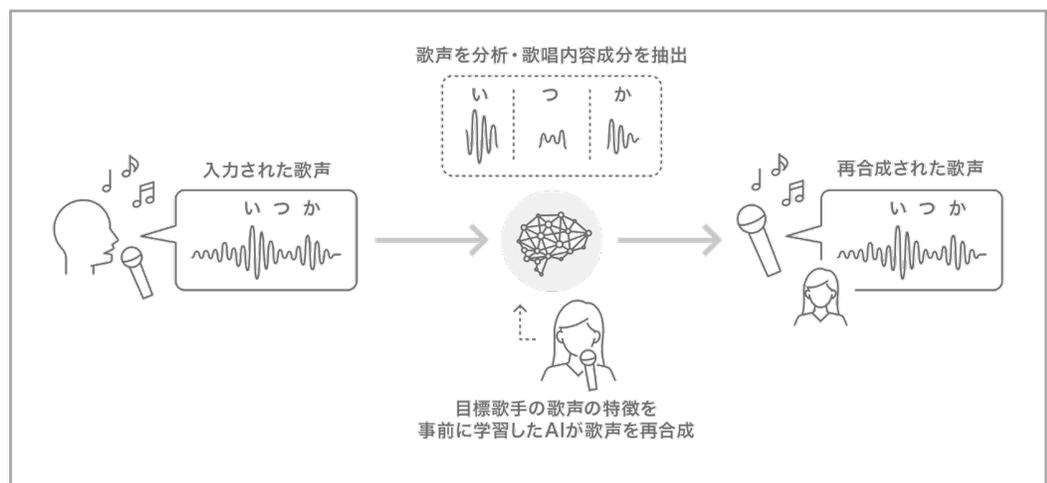


# TransVox

## 技術面



# 「TransVox」と「なりきりマイク」



歌声変換技術「TransVox」



サービス「なりきりマイク」



企画「しゃべる! ふじっぴー」

...

(アーカイブ資料では非公開とさせていただきます)

(アーカイブ資料では非公開とさせていただきます)

## 実地実験の繰り返し

- 研究所（オフィス）で改良 ←→ カラオケ屋さんで実験
- とくにがんばったのは、遅延と耐雑音性
- 「本当の意味で楽しめる」ところまでもっていくことを意識



- 反響

- テレビ取材：たくさん

- スッキリ，ラヴィット，王様のブランチ， etc...

- Web記事：たくさん

- AVWatch, ITMedia, 日経Web, Persolはたわらワイド, etc...

- SNSの反響：たくさん

- 「ELTでいっぱい部屋で、曲の履歴もELTで埋め尽くされて、大好きなもっちゃんの声で歌えて…幸せでした！」

- 振り返り

- 社内デモで仲間を増やすことには、大きな効果があった

- 自分たちだけではこのスピードや規模では進められなかった

# Outline

- 歌声合成技術とは何で、ヤマハは何を目指しているのか
- 具体的な、未来を見据えた歌声合成研究の取り組み
  - 具体例1：VOCALOID  $\beta$ -STUDIO
  - 具体例2：TransVox（なりきりマイク）
- むすび

# むすび

- ヤマハが考える、AIという技術との付き合い方
  - 芸術（音楽）を意思をもって生み出すのは、人間
  - AIと人間が影響しあって何かを生み出すという形を目指す
- 研究からプロダクトへ
  - （まだまだ、本当の製品にはなっていないが…）
  - 利用者のマインドを研究者自身もつこと
    - 才野持論：手がけてる自分自身が、誰より楽しめるか？
  - いろいろな人を巻き込むこと