会誌「デジタルプラクティスコーナー」　Vol.1 No.1(Oct. 2020)

JISA論文

# 声の権利化と流通を実現する音声合成サービス ―一般人から有名人まで多種多様な声が使える新しいプラットフォーム―

金子 祐紀[1]　平林 剛[1]

[1]コエステ（株）

AIやIoTの普及に伴い，音声インタフェースの需要も爆発的に拡大することが予想されている．そこでは，一般人から有名人までさまざまな人の声の合成音声が利用できることが求められると考えられ，我々は東芝の40年以上におよぶ研究開発から生まれた最新の音声合成技術を活用し，一般人から有名人まで多種多様な声を収集・蓄積し，さまざまなサービスで利用可能な音声合成プラットフォームを実現し，提供を開始した．本稿では，プラットフォームの概要とそれを支えるコア技術について述べ，さらに，具体的な導入事例や今後の取り組みについて述べる．

## 1. 声の権利化・流通プラットフォーム「コエステーション」の概要

### 1.1　背景

東芝では40年以上前から，テキストを音声に変換し人工的にさまざまなことを発話させることができる音声合成技術の研究が続けられている．そして，いくつかの文章を読み上げると，その音声からその人の声の特徴を学習し，その人の声の特徴を集めたエッセンスデータである音声合成用声辞書（コエ）を生成，そのコエを音声合成エンジンに適用することで，その人の声に似た合成音声で発話させる技術が生まれた．

これから，AIやIoTの普及に伴い，音声インタフェースの需要が爆発的に拡大するだろうと予想されている．そこでは，「スマートスピーカーの声を好きな芸能人の声にしたい」などの有名人のコエのニーズや，「孫の声で毎朝のニュースが聞きたい」「SNSのメッセージを送信者本人の声で読み上げてほしい」などの身近な一般人のコエのニーズが考えられる．

そこで，上述の技術を活用し，あらかじめ世界中の人々のコエを収集・蓄積し，さまざまなサービスで利用可能な音声合成プラットフォームを実現しようと考えた．

### 1.2　コエステーションの概要

コエステーションは，一般人から有名人まで多種多様なコエを集め，それをさまざまなサービスで活用するための音声合成プラットフォームである．

一般ユーザ向けには，専用のスマートフォンアプリを提供し，ユーザ自身で自分のコエを登録することができるようにしている．アプリをインストールすると，さまざまな文章が表示されるので，それを順番に読み上げていくと，コエステーションのクラウドサーバに読み上げ音声がアップロードされ，ユーザの声の特徴の学習が進み，コエが生成されデータベースに登録される．コエが登録されたら，読み上げ機能で，テキストを入力し音声合成ボタンを押下すると，コエステーションのクラウドサーバで当該ユーザのコエを適用して音声合成処理が行われ，生成された合成音声が再生される．SNS（ソーシャルネットワーキングサービス）で，生成された合成音声を投稿することも可能である．コエの生成に必要なユーザの声の特徴学習のための読み上げ作業は，文章数を多く安定的に読めば読むほど当人の声に似る（似る度合いなどには個人差があるが）．

タレントや声優などの有名人のコエは，ビジネスユースが前提となるため，さらに高品質にする必要があり，通常，防音スタジオで声の特徴学習のための音声収録を行う．約300～400文ほどの文章を読み上げていただき，時間としては1～3時間ほど要するのが一般的である．

このようにして一般人から有名人まで多種多様なコエを収集・蓄積する．そして我々はさまざまなサービス企業に，音声合成ツールとコエの利用権をセットで提供し，ツール利用料とコエ利用料をいただくという事業を展開している．有名人のコエが利用された場合，その声主（コエの権利者）にもコエの利用に見合ったギャランティが支払われるエコシステムを実現している．


図1　コエステーションプラットフォームの概要

東芝デジタルソリューションズ（株）がサービス主体となって，一般ユーザ向けのスマートフォンアプリを2018年4月に，法人向け有償サービスを2018年11月に提供開始した．2020年2月に，エイベックス（株）と東芝デジタルソリューションズ（株）の合弁会社としてコエステ（株）を設立し，現在はコエステ（株）でコエステーション事業は運営されている．

### 1.3　コエステーションで利用可能な法人向け音声合成ツール

コエステーションでは，サービス企業に対し，さまざまなビジネス・サービス形態に適した音声合成ツールを提供している．

1つめは，エディターという，人手で音声合成コンテンツを制作・編集するためのツールである．喜び・怒り・哀しみなどの感情表現や，ピッチ，話速など，簡単な画面操作で細かい作り込みができる．完成した音声は，mp3やwavなどの形式の音声ファイルとしてダウンロードして利用することが可能である．

2つめは，Web APIという，リアルタイムに音声合成処理が必要な場合に利用するツールである．音声チャットボットや毎日情報が変わる天気予報の読み上げなど，あらかじめ音声を作り置き出来ず，都度，音声合成が必要な場合に利用するREST APIで，コエと読み上げさせたいテキストを指定してリクエストすると，コエステーションのクラウドサーバで音声合成処理が行われ，生成された音声データがリアルタイムに取得できる．

もちろん，エディターやWeb APIで生成した合成音声だけでなく，収録音声（生声）と組み合わせてハイブリッドに利用することも可能である．たとえば，大事なキメ台詞だけは収録音声を使うことで臨場感をより向上させるといった使い方が為されている．

通常は，このエディターとWeb APIのツールを利用いただくが，ネットワークに繋がらない環境の場合や，音声合成回数が膨大過ぎてWeb APIではサーバコストが掛かり過ぎてしまうような場合には，組込用ミドルウェアの提供も行っている．

### 1.4　コエの種類

コエステーションで利用できるコエには，有名人などの公式コエと，一般ユーザが専用のスマートフォンアプリで自ら登録した一般コエがあり，コエステーションの音声合成ツールは，コエとセットで利用する．

公式コエの場合，通常，声主は芸能事務所などである．ビジネスユーザから公式コエを利用したいという依頼を受けた場合，我々は声主にその依頼内容を伝え，用途や料金，条件などの調整を行い，折り合いがついたら，その利用に関する契約をビジネスユーザ・声主双方と結び，その契約の範囲でビジネスユーザに利用していただく．コエの利用料に関しては，あらかじめ取り決めたルールに則って，声主にもその一部を支払う．コエの利用料は，用途やビジネスユーザの希望に応じて，月額制や一括支払い，レベニューシェアなど，さまざまな形態がとれるようにしている．

ビジネスユーザが一般コエを利用する場合には，アカウント連携サービス，一般コエ募集サービス，一般コエリクエストサービスのいずれかを利用していただく．

アカウント連携サービスの場合，ビジネスユーザ側のサービスで，コエステーションのIDとパスワードを入力する画面を表示できるように実装していただく．一般ユーザ本人が，ビジネスユーザ側のサービスで，コエステーションと連携するための操作を行い，コエステーションのIDとパスワードを入力し，コエステーション側で認証されると，ビジネスユーザ側のサービスに，当該一般ユーザのコエを利用するためのアクセストークンが発行され，そのコエがそちらのサービスで利用できるようになる．これによって，たとえばSNSのメッセージを送信者本人のコエで読み上げる，プレーヤー本人のコエで喋るキャラクターがゲームで使えるなどといったことが実現可能になる．

また，コエシェア機能によって利用を許可されている家族や友人などのコエもアカウント連携先で利用することができるので，たとえば，恋人のコエで毎朝起こしてくれる目覚ましサービスや，孫のコエで毎日の定時安否確認の連絡をしてくれる高齢者向けサービスなどを実現することが可能になる．

一般コエ募集サービスの場合，コエステーションのスマートフォンアプリで一般ユーザに対してコエ利用を募集することが可能で，応募（許可）された一般ユーザのコエを提示した利用範囲の中で利用することができる．たとえば，朝のニュース番組で毎日違う視聴者のコエを選んで天気情報コーナーのアナウンス音声として使う，アニメーションの新キャラクターのコエを一般から募集するなど，いろいろな活用が考えられる．

不特定多数の一般ユーザに広く募集する一般コエ募集サービスに対して，一般コエリクエストサービスの場合，特定の一般ユーザにコエ利用のリクエストをすることができる．たとえば店長のコエで店内放送を流したいといった場合に，店長があらかじめコエステーションの一般ユーザ向けスマートフォンアプリで自身のコエを登録しておき，一般コエリクエストの承認手続きを踏むことで，エディターやWeb APIといった法人向け音声合成ツールでそのコエを利用することが可能になる．

### 1.5　声の権利

合成音声が生成されるまでにはさまざまな要素が必要である．まず，声優などの声主がコエステーションで提供する専用の文章を読み上げて生声が生成される．その生声をコエ生成エンジンに入力することによりコエが生成される．コエを音声合成エンジンに適用して，読みテキストを入力すると，当該テキストを当該コエで読み上げた合成音声が生成される．読みテキストは，作家が制作した台詞を，ユーザがどう読ませるか（たとえば怒り78%でピッチを少し高くするなど）を調整することで生成される，台詞だけでなくそれをどう読ませるかといったタグ情報も含んだデータである．

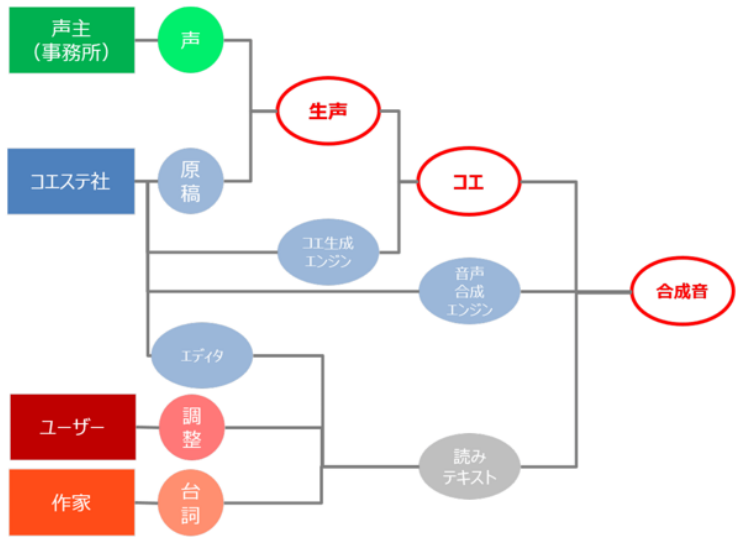


図2　合成音声の生成に必要な要素

我々が弁護士の見解を伺ったところ，まず生声については，用意された短い文章を淡々と読み上げた録音音声データに過ぎず，著作物とは言えないと考えられ，また，その読み上げの際に芸術的な性質を有する演技が行われているわけでもないため，著作隣接権の対象たる実演にも該当しないと考えられるとのことである．

その生声からコエ生成エンジンによりコエが生成されるが，コエは単なるその人の声の特徴を抽出したエッセンスデータであるため，著作物とは言えない．

合成音声は，作家の創作物であるテキストの複製物という扱いになり，作家の著作権が合成音声にも及ぶと考えられる．通常の収録音声の場合，声優などの読み上げた者（実演家）が芸術的な演技を伴って創作されたと見なされた場合には，実演家にも著作隣接権が認められることがあるが，合成音声の場合，声優は合成音声の生成において何も芸術的な演技を行わないため，著作隣接権も認められないと考えられる．尚，ユーザの読み調整により合成音声に創作的な表現が顕れるような場合には，合成音声は作家の創作物のユーザによる翻案物という扱いになり，権利関係は，作家が原著作物であるテキストの著作権を，ユーザが翻案物たる合成音声の著作権を有し，作家は合成音声について原著作者として権利を行使できる，ということになる．

以上により，合成音声およびその生成過程における各種要素について，声主には法的に著作権などは認められない，というのが我々が相談した弁護士から得た見解である．

このため，我々は声主と契約を結ぶことにより，コエが無断で利用されないことや，コエの利用の対価が得られることなどを保証しようと試みており，これがコエステーションを，声の権利化・流通プラットフォームと呼称している所以である．

## 2. プラットフォームの実現を支えるコア技術

### 2.1 多様な声や感情を表現できる音声合成技術

コエステーションでは，隠れマルコフモデル（HMM）という統計モデルに基づく音声合成方式（以下，HMM方式と呼ぶ）を採用している．HMM方式では，音声信号を分析して得られるスペクトルや基本周波数などの音響・韻律パラメータの時系列を統計的にモデル化し，これらを音声合成に用いる．そのため，声質や韻律を，音声波形ではなく，音響・韻律パラメータのレベルで柔軟に操作することが可能で，話者適応などによって多様性を向上させることができる．また，言語への依存性が比較的低く，多言語化も効率的に進めることが可能である[1].

また，合成音声に感情をつけるために，加算構造に基づく感情付与方法を導入している．この技術は，ある目標の話者の平静の音声モデルに対して，事前に複数話者から学習した感情音声と平静音声の差分を表現する感情付与モデルを適用することで，目標話者の感情音声を収録することなく，目標話者の平静音声のみから感情表現可能な音声合成を実現できる．実際に，喜び，怒り，哀しみの3種類の感情を入力文章の任意の部分に任意の強さで付与することが可能である[2].

さらに，「もう少し明るく」や「より女性らしく」といった話者性を表す言葉（知覚表現語）によって，ユーザが直感的に合成音声の声色や口調をデザインできる機能を有している．この機能は，声の特徴を表現する音響・韻律パラメータが，「年齢」「明るさ」などの程度に応じてどのように変化するかを学習したモデル（知覚語空間モデル）を用いて実現している．知覚語空間モデルは，数十名分の学習用話者の音声データと，その音声にどの知覚表現語がどの程度含まれているかを表す得点データとを用いて，各知覚表現語の程度と音響・韻律パラメータとの関係を統計的に学習することであ

らかじめ作成しておく．音声合成時には，ある話者の声のベースとなるモデルからの出力に対して，それぞれの知覚表現語に対応したモデルからの出力を指定の重みを付けて加算することによって音響・韻律パラメータを算出し，所望の声色や口調の音声波形を生成している（**図3**）．法人向けサービスでは，性別，年齢，明瞭さなどの声の特徴を示す11項目のパラメータを調整することで，低コストで簡単にイメージとおりの話者性を持つコエを作り出すことができる[3]．

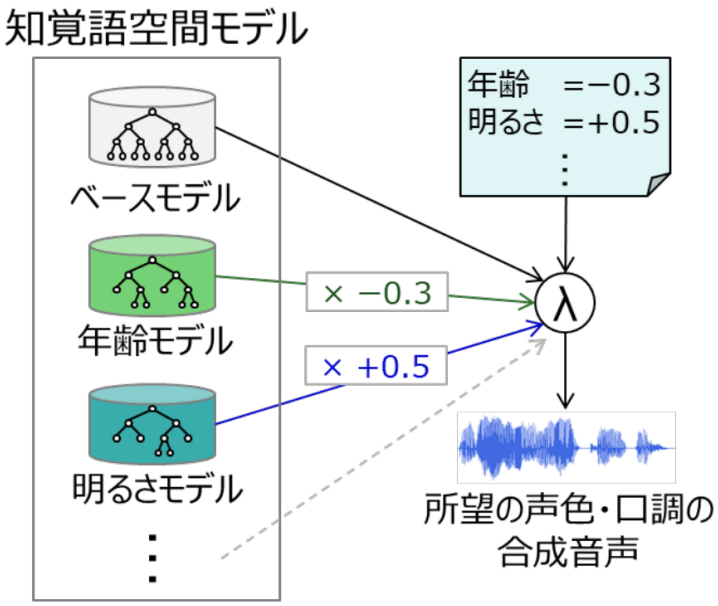


図3　知覚語空間モデルによるコエのデザイン

また，コエの生成に話者適応技術を導入することで，従来の方式よりも少ない録音音声データからでも安定して元話者の特徴を再現することが可能となっている．

具体的には，複数話者の大量の録音音声データから話者に共通な特徴をモデル化したベースモデルをあらかじめ作成しておき，このモデルを目標の話者の特徴に合わせ込むことで，その話者のモデルを生成する．その結果，目標の話者の録音音声データ量が少ない場合でも，言語的な共通の特徴を維持したまま，声質や話し方を録音音声の話者に似たコエを生成することができる．

一般ユーザ向けには，スマートフォンアプリで録音された短い10文章程度の音声データからユーザ自身のコエを作成できるようになっており，録音音声データの解析や話者適応処理などが全自動で実行され，通常であれば 1 時間以内に新たなコエの生成が完了する．

### 2.2　クラウドネイティブなシステム設計

システム全体は，クラウドネイティブに設計しており，Amazon Web Services（AWS）上で稼働している．コエ生成や音声合成など高負荷な処理に対しては，リクエスト数などに応じたオートスケーリングを設定することでリソースの効率的な利用を図っている．さらに，マネージドサービスの活用によって運用・監視コストを抑えることで，一般ユーザ向けサービスの無償提供などを実現している．

### 2.3　なりすまし防止のための電子透かし技術

ユーザ自身のコエが容易に生成可能で，本人に似た合成音声が作成できるようになると，それを「なりすまし」などに悪用される危険性が高まる．そこで，音質への影響が軽微な電子透かしを合成音声に埋め込むとともに，合成音声からこの電子透かしを検出できる技術を開発して採用している．具体的には，人間の聴覚が位相の変化に鈍感であることを利用し，合成音声の位相を緩やかに変化させることで透かしを埋め込んでいる．また検出時は，位相の時系列から変調周波数を求めて判定を行うことができる．この方式は，ノイズや残響，圧縮符号化といった音声の劣化要因の影響を比較的受けにくく，精度の良い検出が可能であるため，有名人や一般ユーザのコエが悪用されることを防ぐのに効果的と考えている．

## 3. コエステーションTMの導入事例

### 3.1 2018 FIFAワールドカップ フジテレビ広報担当「AIカビラくん」

2018年に開催されたサッカーワールドカップで，フジテレビがジョン・カビラ氏をモデルにしたAIキャラクターを制作した．ユーザと対話をしながら，試合の予想をしたりするものであり，コエステーションはその対話の音声発話部分で活用された．

声主としての権利を有するジョン・カビラ氏側には，契約に基づいて対価が支払われた．

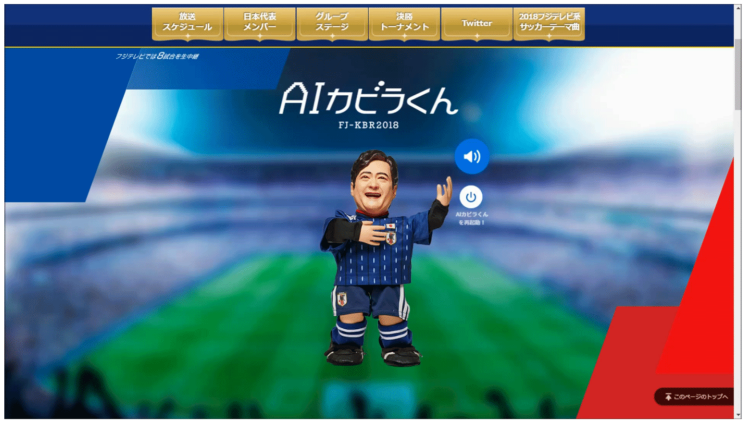


図4　導入事例：AIカビラくん

### 3.2 ウェザーニューズ バーチャルお天気キャスター「WEATHEROID Type A Airi」

天気情報の提供を行っているウェザーニューズが，バーチャルお天気キャスターのオリジナルキャラクターを制作した．Google Homeなどのスマートスピーカーのスキルとして，ユーザと対話をしながら天気情報の提供を行っており，コエステーションはその対話の音声発話部分で活用されている．

本キャラクターのコエの権利はウェザーニューズが保有した形で契約をしているため，声主とユーザが同一であることから，本件における声主対価分は発生しない形で運用されている．

図5　導入事例：バーチャルお天気キャスター「Airi」

### 3.3　モビルス AIアシスタント「バーチャルオペレーター」

　コンタクトセンタソリューションなどを手掛けるモビルスが，AIによって接客や問合せ応答を自動で行うソリューションを開発し，対話の音声発話部分でコエステーションが採用された．

　本事例では，クライアントがコエを選択できる形でのソリューション提供を行っている．タレント・著名人のコエも選択可能で，その際は随時声主との交渉のもと条件をマッチングさせていき，個別契約の範囲でクライアントにコエの利用を許可し，声主側への対価も個別契約に基づいて支払っていく．


図6　導入事例：AIアシスタント「バーチャルオペレーター」

### 3.4　ALS SAVE VOICEプロジェクト

難病ALS患者は病気が進行すると，自力で歩くことや手・指を動かすこと，発話することなどが出来なくなり，表情すら変えることが出来ずに寝たきりになってしまう．そのような患者向けに，視線入力装置の開発・提供を行っているオリィ研究所，ALS患者の支援団体WITH ALSと連携し，視線で文字を入力し，それをコエステーションの音声合成で発話させるシステムを開発した．2019年7月から提供を始めている．

これはアカウント連携サービスによる一般コエを利用した事例である．あくまでユーザ本人（または本件の場合は特別に介助者のケースもあるだろう）による連携操作に基づいて認証されたコエが当該利用範囲に限定して利用できるという仕組みにすることで，なりすましなどの悪用対策をしている．万一，このシステムを使って生成された合成音声が何らかの犯罪などに利用された場合には，当該音声から電子透かしを検出することで，コエステの技術で生成された合成音声かどうかを判別することが可能である．


図7　導入事例：ALS SAVE VOICE PROJECT

## 4. これからの取り組み

### 4.1 グローバル展開とクロスリンガル

現在，コエステーションは日本国内向けのサービスとして展開している．ただ，コエステーションの音声合成技術は11言語に対応しており，近年中にコエステーションはグローバル展開を目指している．

また，研究段階の技術として，クロスリンガルというものがある．現在，日本語の読み上げ音声から学習させて生成するコエでは日本語を発話させることしか出来ないが，クロスリンガル技術により，そのコエで英語や中国語，フランス語などさまざまな言語を発話させることができるようになる．これが実現すると，日本語しか話せない声優の原作の声のまま，アニメや映画を海外展開するなどの用途での活用が考えられる．

### 4.2 デジタルクローンの実現に向けて

生物学的にではなくディジタル技術を使って本人に似たキャラクターをバーチャルに作り上げようといった取り組みも進めている．このような取り組みはデジタルクローンなどと呼ばれている．

人の外見については，毛穴や肌のくすみまで再現する3D CG技術も生まれてきており，脳の思考についてはAIが日々急速に進歩してきている．そして声はコエステーションが利用できるので，さまざまな技術と組み合わせることで，今でもある程度本人に似たデジタルクローンを制作することは可能であり，今後，技術の進歩により，さらにその精度は向上していくと考えられる．

これにより，たとえば本人に代わってデジタルクローンがさまざまな仕事をこなしたり，本人が亡くなってもデジタルクローンが家族の相談に乗ったりといったことが可能になるかもしれない．

## 5. おわりに

音声認識・音声合成関連の世界市場は，2025年には約20兆円規模まで拡大するといった予測（出典：BCCリサーチ社音声認識・合成技術関連の世界市場規模）も出ているように，現在，音声インタフェースの業界は大きな注目を集めており，実際，多くの喋る機器や対話ベースの業務ソリューションなどが出てきている．しかし，そこで選べる声はせいぜい数種類程度であり，声は固定で切り替えられないものの方が未だに多い．

ただ，ユーザニーズを伺うと，「カーナビの声を自分の好きな芸能人の声にできるなら課金してでも使いたい」「オーディオブックやニュース読み上げアプリなどを好きな声で聴きたい」といった声は非常に多く，音声インタフェースで声が選べるというのは今後あたり前になってくるだろうと考えている．そして，有名人のコエだけでなく，SNSのメッセージを本人のコエで読み上げるなど，一般ユーザのコエのニーズも今後高まっていくだろうと予想している．

では，なぜ人々は，好きな芸能人の声で聴きたいとか，本人の声で読み上げてほしいと考えるのだろうか．それは，声が非常にエモーショナルなものであり，人のアイデンティティにおいて重要な要素だからである．実際，認知症患者が家族の声にだけ反応するといった事例も多い．

コエステーションは，人間の発話するという機能をディジタル保存し，それを手軽に利用できるようにした世界初のプラットフォームである．世界70億人のコエを，安心してさまざまなサービスで活用できる，100年後にも役立つディジタルインフラの実現が，コエステーションの目標である．

## 参考文献

1）森田眞弘 他：多様な声や感情を豊かに表現できる音声合成技術．東芝レビュー Vol.68, pp.10-13 (Sep. 2013).
2）大谷大和 他：HMMに基づく感情音声合成のための共有感情付与モデル．信学技報 SP2014-92, pp.13-18 (Nov. 2014).
3）東芝：話者の声の特徴を直感的な言葉で制御できる音声合成技術，東芝レビュー Vol.71, pp.80-83 (Apr. 2016).

**金子 祐紀**（非会員）kaneko-yuki@av.avex.co.jp
2005年に東芝に入社し，クラウドTVやメガネ型ウェアラブルなど, さまざまな新規事業の立ち上げに携わる社内起業家．2016年から音声合成技術を活用した声の新しいプラットフォーム「コエステーション」を立ち上げ, 2020年2月にエイベックスとの合弁会社であるコエステ（株）を設立. 2020年4月にコエステ（株）の執行役員に就任. 東京大学大学院協力研究員.

**平林 剛**（非会員）hirabayashi-go@av.avex.co.jp
1998年に東芝入社．研究開発センタにて音声合成技術の研究・開発に従事．2010年からは音声合成技術の事業化にも取り組み，企業向けクラウドサービスやコンシューマー向けWebサービスの企画・運営などを担当．2020年2月にエイベックスとの合弁会社であるコエステ（株）を設立. 2020年4月にコエステ（株）の技術統括責任者に就任..

採録決定：2020年7月14日
編集担当：細野 繁（東京工科大学）