



# AI音声合成の技術動向

高道 慎之介(東京大学)



@forthshinji

## 名前

高道 慎之介 (たかみち しんのすけ)

## 現職

東京大学 講師

## 経歴

熊本高専→長岡技大→奈良先端大.  
2016年に博士 (工学).

## 専門

音声処理 (主に音声合成)

# 最近のAI音声合成は本当にスゴイ！



音声合成

× 🔊 🔍

くすべて 動画 ニュース 画像 ショッピング もっと見る ツール

約 813 件 (0.27 秒)

**ASPIC**  
音声合成ソフト比較10選。法人向けの機能やタイプ、選び方は？ | アスピック  
音声合成ソフトとは、テキストをもとに人工的な音声を作り出し、読み上げてくれるソフトです。従来、機器や動画コンテンツにナレーションを入れるに...  
6日前

**MdN Design Interactive**  
AI音声合成を活用したオーディオブックアプリ登場、池澤春菜さんなど人気声優のAI音声でニュースや書籍を読み上げる  
株式会社GADGETは7月25日、AI音声合成を利用したオーディオブックアプリ「VICKE Audiobook&News (ビッケ オーディオブックアンドニュース)」の提供...  
4日前

**PR TIMES**  
日本初AI音声合成によるオーディオブックアプリ7月25日開始  
有名声優の声による音声合成事業を展開する株式会社GADGET (本社:東京都渋谷区、代表取締役:浅見敬、以下ガジェット)は、HOYA株式会社 (本社:東京...  
1週間前

**RS Real Sound | リアルサウンド**  
愛美&相羽あいなの声による音声合成ソフト「夢ノ結唱」シリーズが与える新体験 ナナホシ管弦楽団、雄之助らの制作楽曲を聴いて  
ChatGPTなどの対話型AIや画像生成AIツールなどが話題となっている昨今。音楽の分野においても、音楽そのものを自動生成するツールが数多く登場して...  
3週間前

Google ニュース検索 (2023年8月2日 時点)

- 最近のニュースを見てみると…
- 新しいボイスの登場
  - “声優〇〇の音声合成が登場！”
  - “シリーズ△△の新キャラクタが登場！”
- 新しいサービスの登場
  - オーディオブック
  - Podcast
  - **ゲーム ← どんな技術が研究されており、どんな権利を検討すべきだろうか？**

AI音声合成とは、人工知能技術(AI技術)を用いて音声を合成する技術です。

本講演では、大学研究者である私(高道)から音声合成の技術動向について、  
**前半の内容**

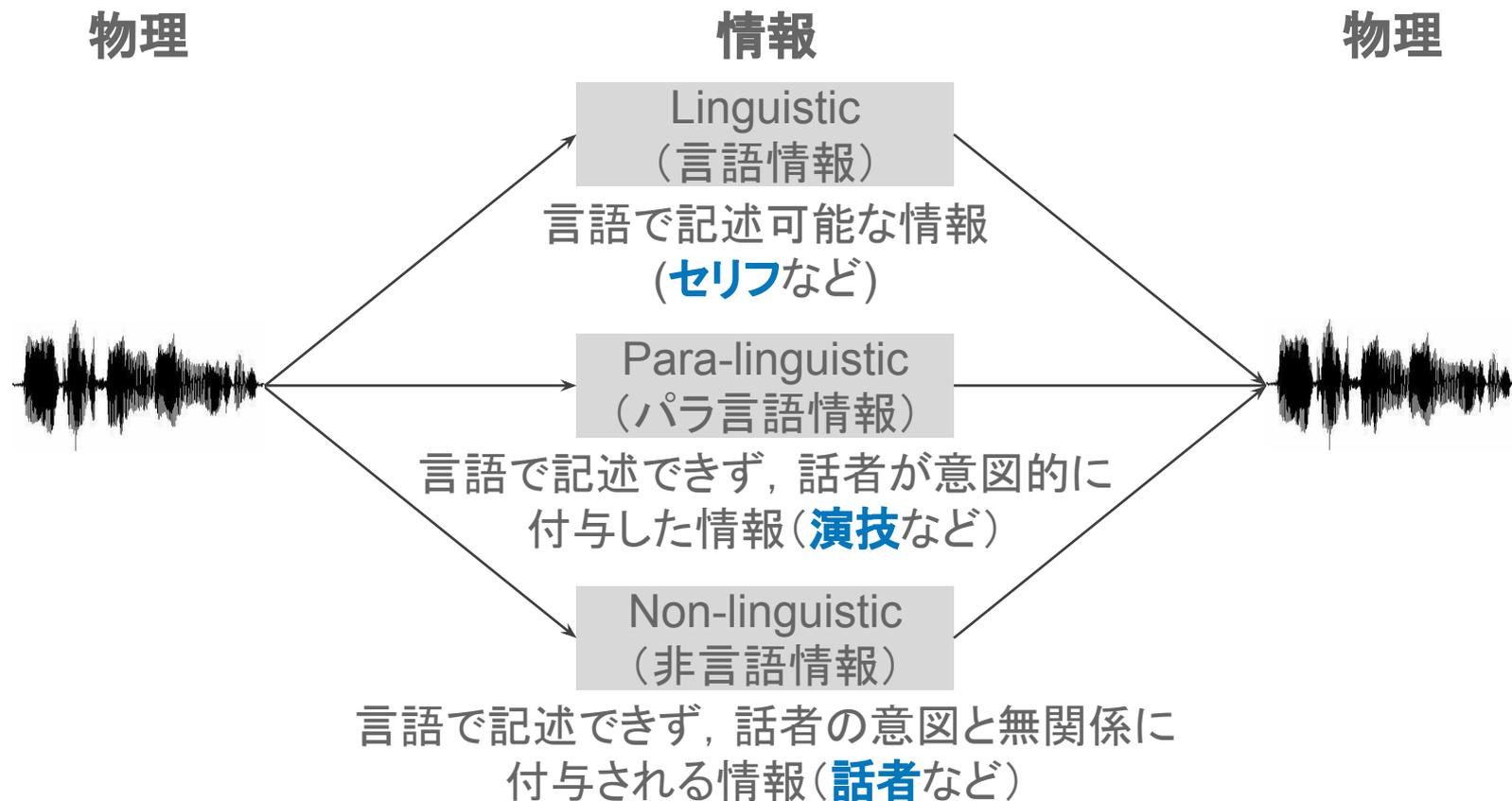
共同研究者である柿沼太一弁護士(STORIA法律事務所)ご協力の元、音声合成の  
**後半の内容**

権利について、それぞれ紹介します。

本講演では、運営側から事前にいくつか質問を頂いています。  
前半と後半それぞれで回答(検討)していきます。

# 最近の技術動向

# 前提知識: 音声の持つ情報



# 基礎研究の段階を終え，開発・実装段階に進んだもの



- 言語情報: 音声資源の豊富な言語における，高品質なテキスト読み上げ
  - 世界の7000言語のうち，数十言語が該当
  - それ以外の言語(方言，希少言語，消滅言語，人工言語)はこれから
- パラ言語情報: カテゴリカルな感情の表現
  - 例えば喜怒哀楽，およびその強弱
  - 複雑な表現(曖昧指示，環境，文脈に起因するもの)はこれから
- 非言語情報: カテゴリカルな話者の表現
  - 例えば1000人の話者の切り替え
  - 存在しない話者・話者属性などはこれから

# 非言語音声の確率的生成: テキストで書けないセリフの音声を合成しよう

”こんにちは” — AI音声合成 →   
文字で書ける音声なら音声を合成できそう!

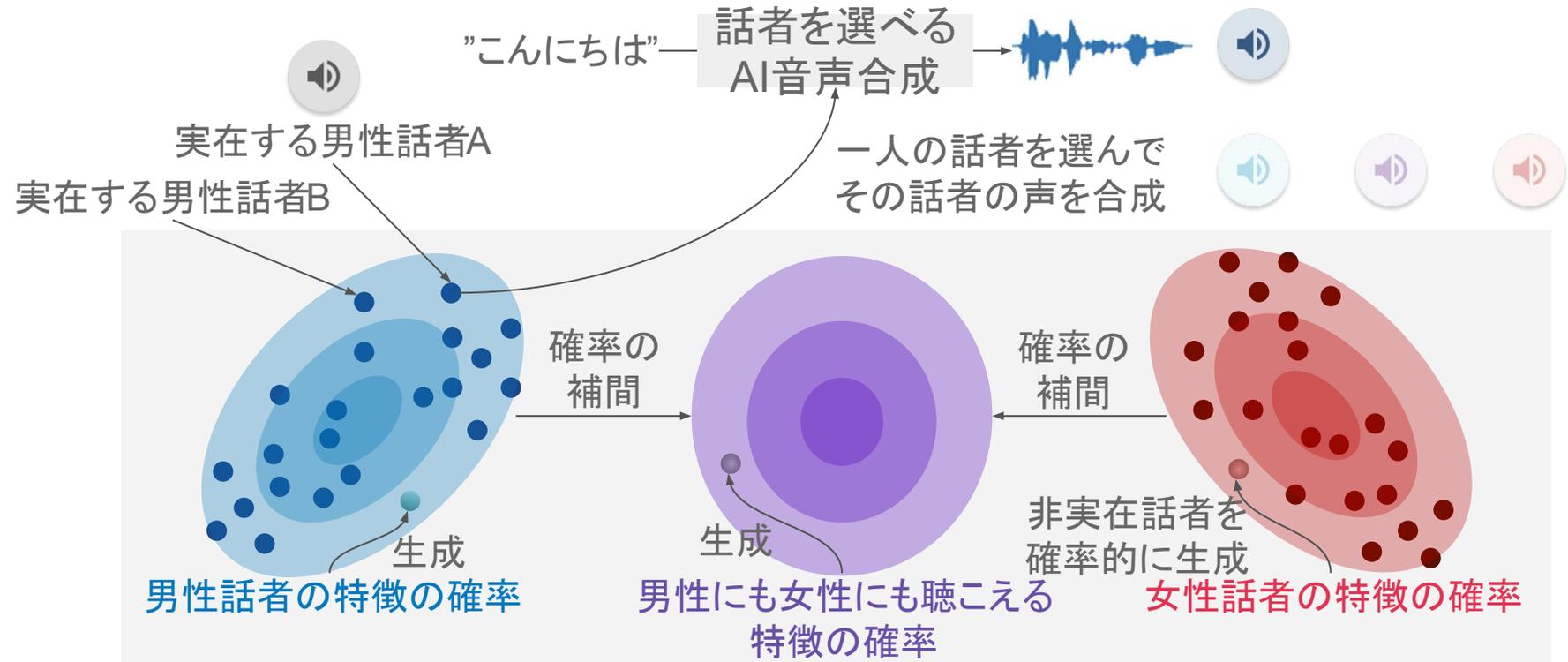
”笑” or “www” — AI音声合成 → ???  
文字で書けない音声をどうやって合成する?

- 例えば笑い声の生成



Xin: “Laughter Synthesis using Pseudo Phonetic Tokens with a Large-scale In-the-wild Laughter Corpus,” Interspeech, 2023.

# 非実在話者の生成: 実在しない話者の声で喋るAI音声合成を作ろう

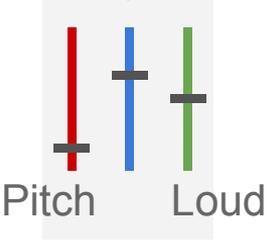
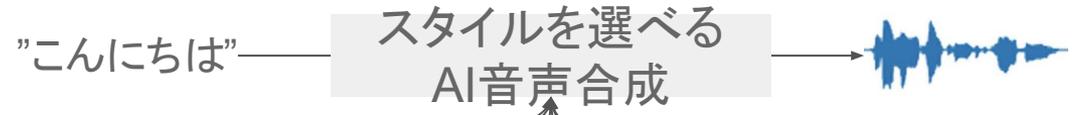


話者の特徴を表す数値(例えば, 音高, 声色を表す多次元の数値)

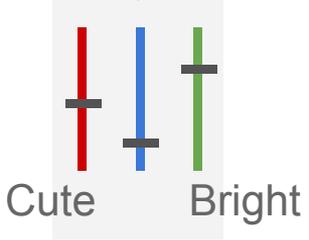
Stanton: "Speaker generation," ICASSP 2022.

Watanabe: "Mid-attribute speaker generation using optimal-transport-based interpolation of Gaussian mixture models," ICASSP 2023.

# 非実在スタイルの生成: 実在しないスタイルの声で喋るAI音声合成を作ろう



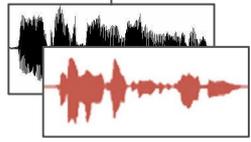
**特徴量**  
音声特徴量を直接制御



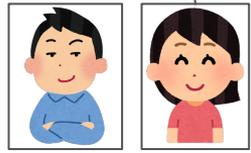
**カテゴリ**  
声質カテゴリを制御



**参照音声**  
音声の声質やスタイルを模倣



**参照&補間**  
参照の声質やスタイルを補間



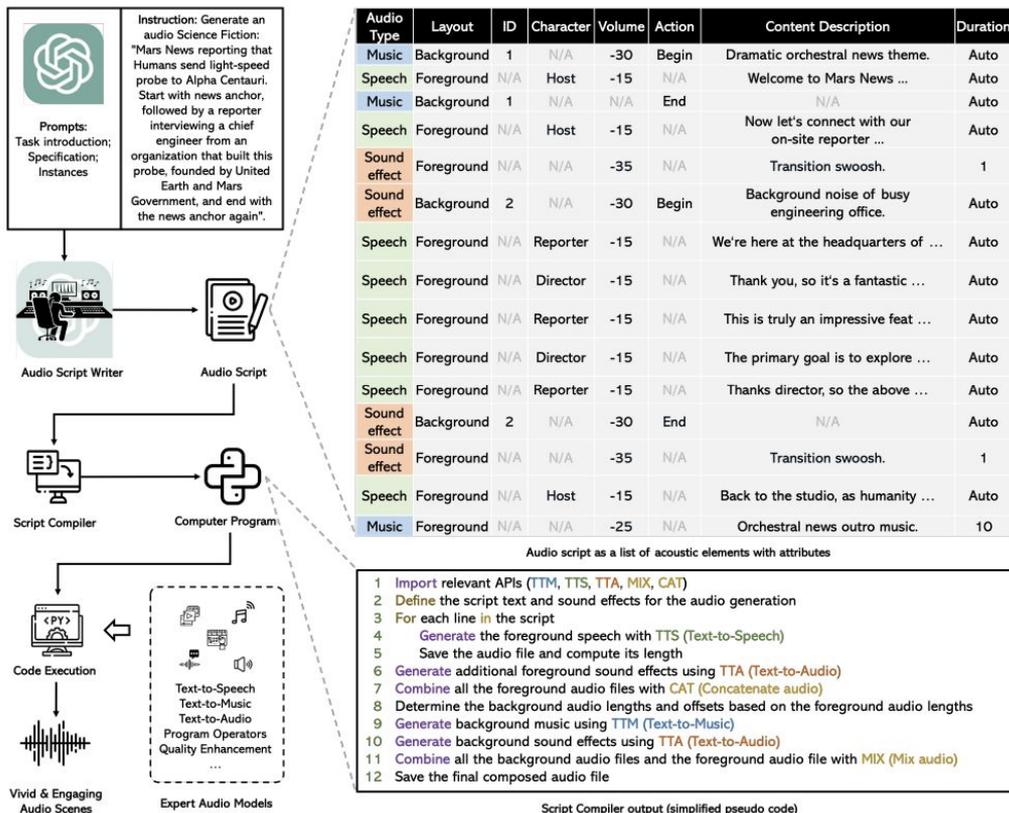
**参照(動)画像**  
(顔)画像に対応する声質を推定

”20代男性が快活に喋る声”

**テキスト**  
文から声質やスタイルを推定  
[近年登場]

Guo: "PromptTTS: Controllable Text-to-Speech with Text Descriptions," ICASSP 2023.  
渡邊: "Coco-Nut: 自由記述文による声質制御に向けた多話者音声・声質自由記述ペアデータセット" 音響学会 (秋), 2023.

# マルチトラックの音生成: 音声・音イベント・音楽を構造的に生成しよう



- 大規模言語モデル(例えば ChatGPT)を使った音プロンプトの生成
- 環境音, 音楽, 音声のスクリプトを自動作成 → 音合成
- 論文中でも言及されているとおり, 音合成の品質が現時点で十分ではないが, システムコンセプトとしては期待

Liu: "WavJourney: Compositional Audio Creation with Large Language Models," arXiv 2023.

# AI音(声)合成技術に求められていることは何だろうか？

”お湯をコップに注ぐ音”

音合成



- 音声合成に関連する技術として音合成(音イベントなどを合成する技術)がある
  - 2023年に、映画音響エンジニアに対する技術サーベイが行われた。
  - これらの議論は、音声合成でも検討すべきだろう。
- Q1: 音の収録における課題はなんですか？
  - A1: 映像などとの時間同期, 音質, 他録音との音色の一貫性
- Q2: 現在の text-to-audio (上図) の問題点はなんですか？
  - A2: 音質・忠実度の低さ, 粗さ. 創造性を満たせる音生成の方法論
- Q3: 何から音を合成したいですか？
  - A3: 時間同期させる映像, 参照音, 参照補間

Oh: "A Demand-Driven Perspective on Generative Audio AI," arXiv 2023.

# 創作のための音声合成AI

- 個人的な見解
  - 現時点で、人間による音声創作を代替・拡張しうる技術は非常に限定的ではないだろうか(少なくとも、予想できる範囲を大きく超えてはいない).
  - これは、今の音声合成AIが従来の「テキストを高品質に読み上げる+ $\alpha$ 」の域を超えていないから？
- 今後はどう進むだろうか？
  - 音声合成AIが「テキストを高品質に読み上げる+ $\alpha$ 」を超えたときに大きな変化があると思われる。現時点でその片鱗を見せる技術はいくつか存在する。
  - まずは、テキストで書けない音声の生成(自発性を含む)、カテゴリカルでないパラ言語・非言語の生成が進んでいくのではないだろうか。
  - そのあとに、創作のための音声AIとしての、UI/UXデザインの研究が進む？

# 個人的に重要なキーワードは「自律」

- 行動の自律性
  - 環境と相互作用しつつ、初期の設計や行動時の持続的な介入によることなく動作する能力や、さらに その上で目標を作り達成する能力
- ゲーム内のAIキャラクターが「行動の自律性」をもつ音声合成とは？
  - 環境(広くは世界)に作用されて喋る
  - 他者(人間, 他AI)に作用されて喋る
  - 自身に作用されて喋る
  - → AIキャラクターに限らず, 何らかの外的要因に自律的に作用される, あるいは外的要因に作用する機能までを備えた音(声)生成が必要だと思われる.

# 質問:収録の仕事は減るのだろうか？



- 設備について
  - 収録スタジオは、音響的な物理空間としての役割よりも、コミュニケーション空間としての役割が重要になるかも。
  - 人間と人間、人間とAI、あるいは人間と環境のコミュニケーションをどうデザインするか？
- 人間について
  - 演技の方法論において本質的なところは依然として変わらないのではないかと(私は演技の専門家ではありませんが…)
  - 音声合成AIは、何らかの制約(例えば、音声を司る器官)から人間を開放するための技術
  - ただし、技術、倫理、法律、産業の専門家を集めた早めの議論は必要だと思われる。

# 質問: 違和感のある合成音声を修正するには？



- 根本的な問題は「合成音声AIは自分の出力する音声を聞いていない」ことだと思われる。
  - 人間は、自分の音声をリアルタイムにモニタリングする機構を有する(“言葉の鎖”)
  - まずは、この機構を計算機的に実現することが目下の課題
- そのあとに、人間とコミュニケーションするためのUIデザインが成熟していく？
  - 今でもいくつか研究されている(例えば、異なるピッチパターンの生成、部分的なセリフ置換などなど)
  - 音声合成AIが「テキストを高品質に読み上げる+ $\alpha$ 」を超えた後に、急速に進んでいくのではないかと思われる。

## 質問: 今後, 創作のための音声合成AIの研究は どう進んでいくだろうか?

- 「テキストを高品質に読み上げる+ $\alpha$ 」を超えるための技術・データベース
  - 特に, 技術発達と速度と比較して, 機械学習のための音(声)データベースの整備が遅れているのは事実.
- 大規模言語モデル(例えばChatGPT)との接続, 環境からの作用, 環境への作用
- 「テキストを高品質に読み上げる+ $\alpha$ 」を超えた音声合成AIの早急な展開
  - 音声UXのデザインをあつかう人に, 早く手渡せるように

# 音声合成の権利

# 柿沼太一先生ご紹介



- 1997年京都大学卒業
- 2000年4月弁護士登録
- 2015年STORIA法律事務所設立
- AI、IT、知的財産、スタートアップ法務を主として取り扱う
- 経産省の「AI・データ契約ガイドライン」検討委員会委員 (2018)
- スタートアップファクトリー構築事業に係る契約ガイドライン検討会構成員 (<https://startup-f.jp/>) (2018)
- 経済産業省「オープンイノベーションを促進するための技術分野別契約ガイドラインに関する調査研究」委員会事務局(2020～)

# 1 設例

- (1) 様々なゲーム内の、様々な声優による音声を大量に収集し、音声生成用AIを作成したい。無断でそれらの声優音声データを収集してAIの生成(学習)に利用することは可能か。
- (2) ある特定の声優の音声を再現するために、当該声優の音声を大量に収集し、音声生成用AIを作成したい。無断でそれらの声優音声データを収集してAIの生成(学習)に利用することは可能か。
- (3) 作成した音声生成用AIを利用して、①特定のゲーム内の特定の声優による特定のセリフ音声と同じ音声を生成すること、または②特定の声優による、それまでに実演されたことのないオリジナルセリフ音声を生成することは可能か。

以降のスライドで、これらの設例を順に検討する

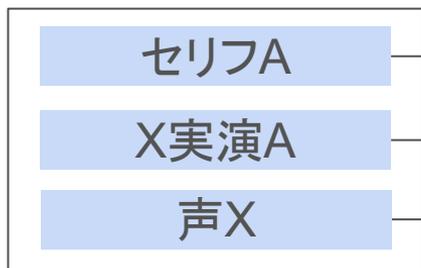
# 1 設例: 基本的な考え方



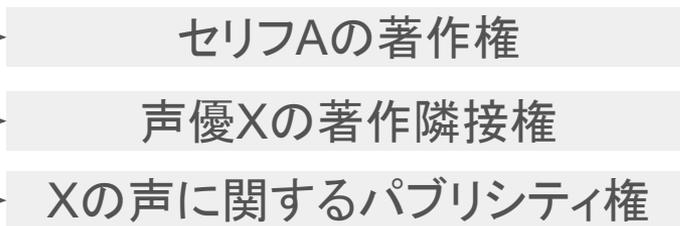
- 以下の2点について検討する必要がある。
  - (1)ゲーム内の声優の音声に関して誰がどのような権利を持っているか
  - (2)音声生成AIの作成(学習)フェーズと、音声生成AIによって生成された実際の音声の利用フェーズそれぞれにおいて、(1)の権利の侵害行為があるか

## 2 ゲーム内声優音声に関する権利関係

声優XによるセリフAの実演



声優XによるセリフAの実演に関する各種権利



- ゲーム内声優音声に関する権利関係としては、①セリフの著作権、②声優の実演に関する著作隣接権（セリフという著作物を実演することにより生じる権利）、③声優の声に関するパブリシティ権が存在する。
  - ①セリフの著作権: セリフを作成した著作者（脚本家等）に生じる。
  - ②声優の実演に関する著作隣接権: 実演をした声優という実演家に生じる。
  - ③声優の声に関するパブリシティ権: 声優に生じる。

## 2 ゲーム内声優音声に関する権利関係： 声優やアーティストの「声」はパブリシティ権で保護されるのか

### 1 ピンク・レディー事件(最高裁平成24年2月2日判決)

- 人の氏名, 肖像等(以下, 併せて「**肖像等**」という。)は, 個人の人格の象徴であるから, 当該個人は, 人格権に由来するものとして, これをみだりに利用されない権利を有すると解される(略)。
- そして, **肖像等は, 商品の販売等を促進する顧客吸引力を有する場合があります, このような顧客吸引力を排他的に利用する権利(以下「パブリシティ権」という。)**は, 肖像等それ自体の商業的価値に基づくものであるから, 上記の人格権に由来する権利の一内容を構成するものといえることができる。
- 他方, 肖像等に顧客吸引力を有する者は, 社会の耳目を集めるなどして, その肖像等を時事報道, 論説, 創作物等に使用されることもあるのであって, その使用を正当な表現行為等として受忍すべき場合もあるというべきである。
- そうすると, 肖像等を無断で使用する行為は,
  - ①肖像等それ自体を独立して鑑賞の対象となる商品等として使用し,
  - ②商品等の差別化を図る目的で肖像等を商品等に付し,
  - ③肖像等を商品等の広告として使用するなど,
- 専ら肖像等の有する顧客吸引力の利用を目的とするといえる場合に, パブリシティ権を侵害するものとして, 不法行為法上違法となると解するのが相当である。

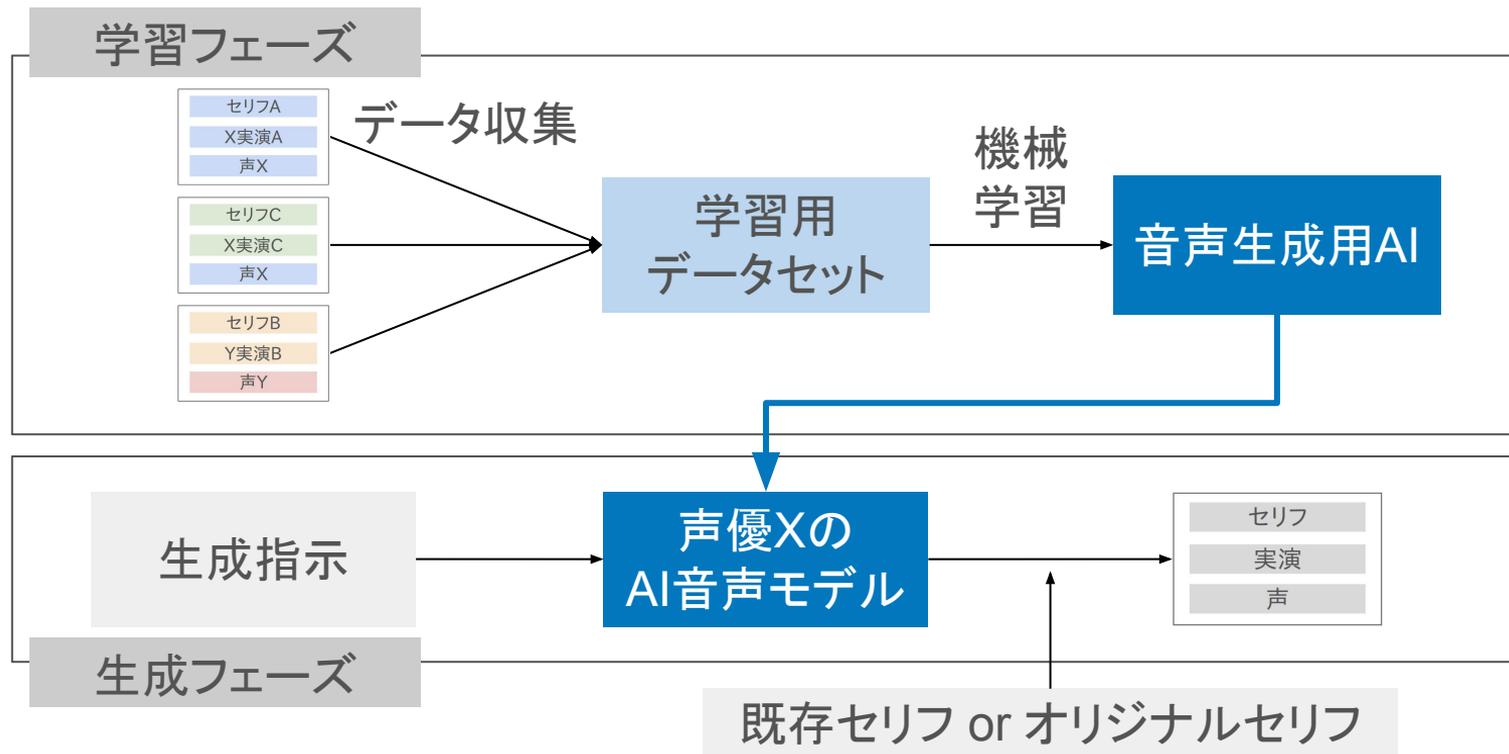
## 2 ゲーム内声優音声に関する権利関係： 声優やアーティストの「声」はパブリシティ権で保護されるのか

### 2 ピンク・レディー事件の調査官解説(民事篇平成24年度(上)18頁)

- 「本判決の3類型にいう「肖像等」とは、本人の人物識別情報をいうものであり、たとえば、サイン、署名、**声**、ペンネーム、芸名等を含むものである。」
- →したがって、声優の「声」を以下の態様で利用する行為は、声優のパブリシティ権を侵害する。
  - ①声それ自体を独立して鑑賞の対象となる商品等として使用し、
  - ②商品等の差別化を図る目的で声を商品等に付し、
  - ③声を商品等の広告として使用するなど、
- 専ら声の有する顧客吸引力の利用を目的とするといえる場合

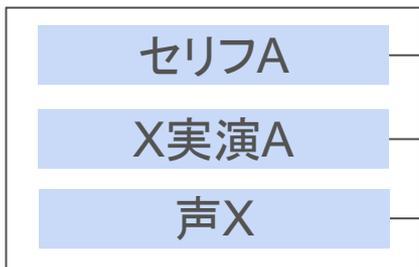
→当該「声」が当該声優の声であると認識できれば「何をしゃべっているか」とは無関係にパブリシティ権が発生し、当該声優に無断で当該声を上記態様で利用すれば、パブリシティ権侵害になる。

# 3 学習フェーズと生成フェーズ

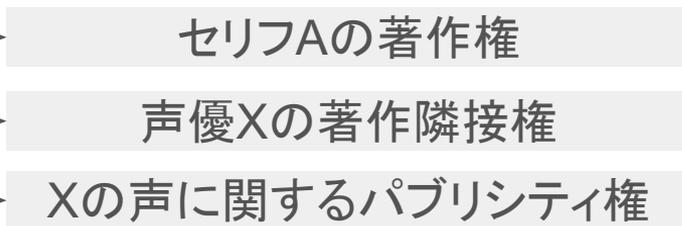


# 4 学習フェーズの問題

声優XによるセリフAの実演



声優XによるセリフAの実演に関する各種権利



## 1. セリフAの著作権

- a. **著作権法30条の4第2号**により、セリフAの著作権者の許諾なく学習に利用しても原則として適法。

## 2. 実演家Xの著作隣接権

- a. 著作権法第102条、同30条の4第2号により、実演家Xの許諾なく学習に利用しても原則として適法。

## 3. Xの声に関するパブリシティ権

- a. 著作権法の問題ではない。学習に利用するだけであれば、ピンク・レディー事件が掲げた「専ら肖像等の有する顧客吸引力の利用を目的とするといえる場合」に該当せず適法ではないか。

## 4 学習フェーズの問題

### 【著作権法】第三十条の四

著作物は、次に掲げる場合その他の当該著作物に表現された思想又は感情を自ら享受し又は他人に享受させることを目的としない場合には、その必要と認められる限度において、いずれの方法によるかを問わず、利用することができる。ただし、当該著作物の種類及び用途並びに当該利用の態様に照らし著作権者の利益を不当に害することとなる場合は、この限りでない。

一(略)

二 情報解析(多数の著作物その他の大量の情報から、当該情報を構成する**言語、音、**映像その他の要素に係る情報を抽出し、比較、分類その他の解析を行うことをいう。第四十七条の五第一項第二号において同じ。)の用に供する場合

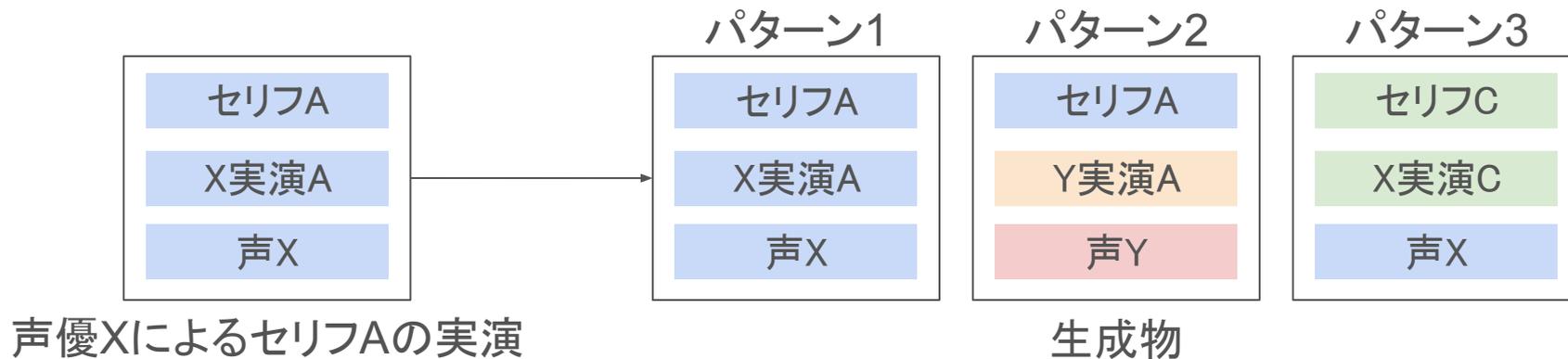
三(略)

## 4 学習フェーズの問題

- (1) 様々なゲーム内の、様々な声優による音声を大量に収集し、音声生成用AIを作成したい。無断でそれらの声優音声データを収集してAIの生成(学習)に利用することは可能か。→ **【検討】適法**
- (2) ある特定の声優の音声を再現するために、当該声優の音声を大量に収集し、音声生成用AIを作成したい。無断でそれらの声優音声データを収集してAIの生成(学習)に利用することは可能か。→ **【検討】適法**
- (3) 作成した音声生成用AIを利用して、①特定のゲーム内の特定の声優による特定のセリフ音声と同じ音声を生成すること、または②特定の声優による、それまでに実演されたことのないオリジナルセリフ音声を生成することは可能か。

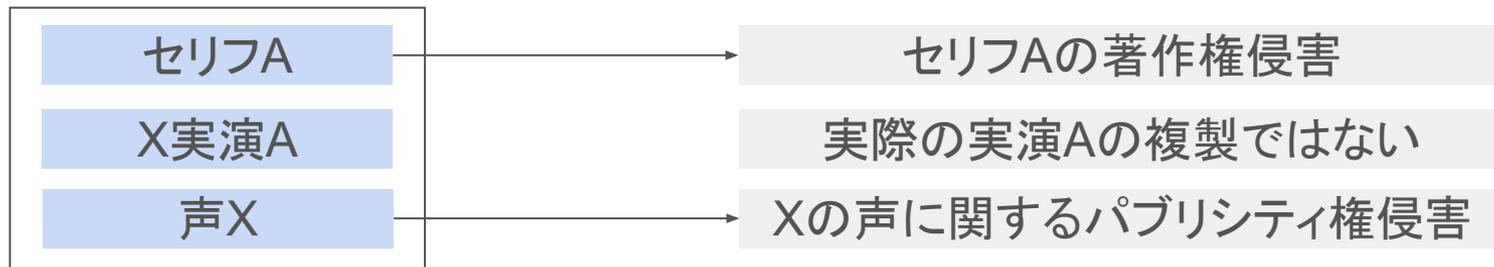
# 5 生成フェーズの問題

- 生成物の3パターン
  - 1 既存セリフ+実演+声
    - ① オリジナルと同じ声優Xによる既存セリフAの実演:パターン1
    - ② オリジナルと異なる声優Yによる既存セリフAの実演:パターン2
  - 2 オリジナルセリフ+実演+声
    - 声優Xによるオリジナルセリフの実演:パターン3



## 5 生成フェーズの問題

### (1) パターン1



## 5 生成フェーズの問題

### (1) パターン1

- 問題

- 音声生成AIを利用して、既存セリフAの声優Xの実演を作成してゲームに利用した場合、誰のどのような権利を侵害するのか。

- 検討

- (1) セリフの著作権

- 単なる既存セリフの無断利用として**著作権侵害になる**(生成AIの問題ではない)。

- (2) 声優Xの著作隣接権

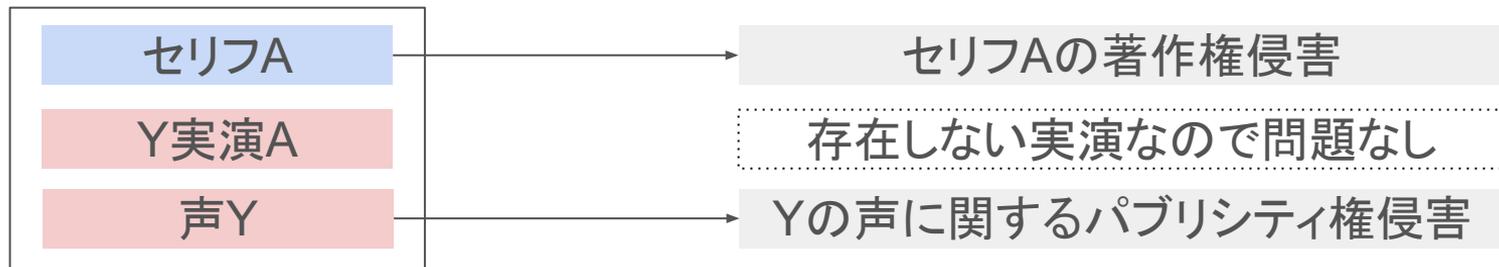
- 実演家は、「その実演」(実際に演じられた実演そのもの)に関する隣接権を有する(法第1条等)が、AIにより生成・出力された実演は「その実演」とはいえない。
- したがって、**声優Xの著作隣接権の侵害にはならない**。
- \*「(引用者注:実演家の録音権)は実演家の実演そのものを録音・録画する権利であるから、たとえば、ある歌手のもの真似をする行為は録音・録画には該当せず、実演家の隣接権侵害にはならない」(中山信弘「著作権法(第3版)」P668)

- (3) 声優Xのパブリシティ権

- 音声生成AIにより生成された音声であっても、当該「声」が当該声優の声であると認識できれば「何をしゃべっているか」とは無関係にパブリシティ権が発生し、当該声優に無断で当該声をゲーム内音声として利用すれば、**声優Xのパブリシティ権侵害になる**。

## 5 生成フェーズの問題

### (2) パターン2



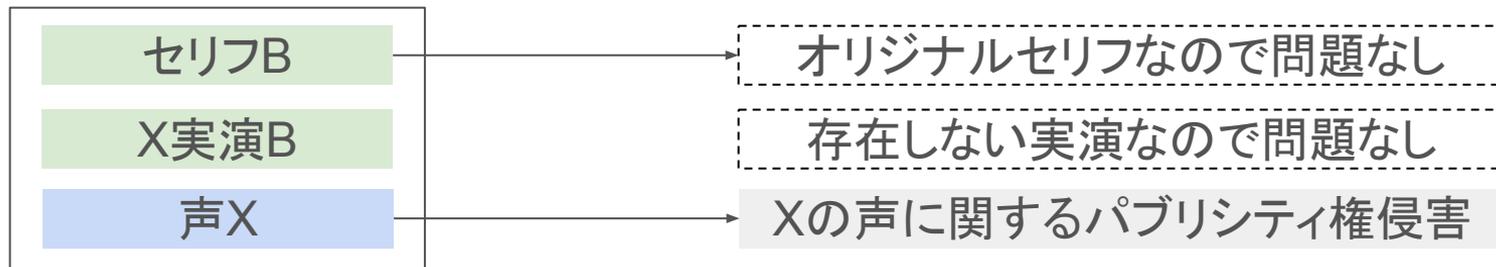
## 5 生成フェーズの問題

### (2) パターン2

- 問題
  - 音声生成AIを利用して、既存セリフAについて、オリジナル声優Xではない、声優Yの実演を作成してゲームに利用した場合、誰のどのような権利を侵害するのか。
- 検討
  - (1) セリフの著作権
    - 単なる既存セリフの無断利用として**著作権侵害になる**(生成AIの問題ではない)。
  - (2) 声優Yの著作隣接権
    - 声優YによるセリフAの実演は存在しない実演なので、**声優Yの著作隣接権の侵害にはならない**。
  - (3) 声優Yのパブリシティ権
    - 音声生成AIにより生成された音声であっても、当該「声」が当該声優の声であると認識できれば「何をしゃべっているか」とは無関係にパブリシティ権が発生し、当該声優に無断で当該声をゲーム内音声として利用すれば、**声優Yのパブリシティ権侵害になる**。

## 5 生成フェーズの問題

### (3) パターン3



## 5 生成フェーズの問題

### (3) パターン3

- 問題
  - 音声生成AIを利用して、オリジナルセリフAについて、声優Xの実演を作成してゲームに利用した場合、誰のどのような権利を侵害するのか。
- 検討
  - (1) セリフの著作権
    - オリジナルセリフの利用なので、著作権侵害にはならない。
  - (2) 声優Xの著作隣接権
    - 声優Xによるオリジナルセリフの実演は存在しない実演なので、声優Xの著作隣接権の侵害にはならない。
  - (3) 声優Xのパブリシティ権
    - 音声生成AIにより生成された音声であっても、当該「声」が当該声優の声であると認識できれば「何をしゃべっているか」とは無関係にパブリシティ権が発生し、当該声優に無断で当該声をゲーム内音声として利用すれば、声優Xのパブリシティ権侵害になる。

## 5 生成フェーズの問題

(1) 様々なゲーム内の、様々な声優による音声を大量に収集し、音声生成用AIを作

### 【検討】①について

- セリフについての著作権侵害に該当。
- 声優Xの著作隣接権は侵害しない。
- 声優Xのパブリシティ権侵害に該当。

### 【検討】②について

- セリフについての著作権侵害には該当しない。
- 声優Xの著作隣接権は侵害しない。
- 声優Xのパブリシティ権侵害に該当。

(3) 作成した音声生成用AIを利用して、①特定のゲーム内の特定の声優による特定のセリフ音声と同じ音声を生成すること、または②特定の声優による、それまでに実演されたことのないオリジナルセリフ音声を生成することは可能か。

## 6 いくつかの具体的検討

- 設問

- 画像生成AIで倫理の問題でヨーロッパで法律ができるかという話を聞きますが、音声合成では法規制はどうなっていくと予測しますか？（日本に限らず世界的な動向として）

- 検討

- 音声合成特有の規制はされないだろうと予想。
- ただし、「特定の人物Aの音声」は「特定の人物A」と強く結びついている
  - →虚偽の内容の音声や、当該人物Aの名誉や信用を毀損するような態様での利用（例：アダルト的な利用や人物Aの思想信条と異なる内容の音声）は、**パブリシティ権侵害とは別に名誉毀損などに該当する可能性が高い。**
  - →そのようなケースについては、不法行為による損害賠償請求や、場合によっては刑事罰（名誉毀損等）で対応することになる。

## 6 いくつかの具体的検討

- 設問

- 既に収録済みの音声をイントネーションや抑揚の調整を行った場合、声優本人の権利はどう扱われるべきでしょうか？ただし、声優の声を使うのにあたりその人の印象が悪くなるような使い方をしないものとします。
  - →これを行う事により、同意がとれていない台詞を後から作って実装できてしまうため、音声合成の話に置き換えられる可能性もありそうだと感じている。

- 検討

- 既に収録済みの音声をイントネーションや抑揚の調整を行うに過ぎない場合は、その点について声優の黙示の同意があるとみなされる場合は多そう。そのような場合であれば権利侵害の問題は生じない。
- 一方、既に収録済みの音声をを用いてAIを作成し、同意がとれていないセリフを後から作って実装することは、黙示の同意があるとは言えないため、パブリシティ権侵害に該当すると思われる。

## 6 いくつかの具体的検討

- 設問

- 音声合成の完璧なモデルが出来てしまえばその声優さんの収録の必要が無くなりそうです。声優の収録で発生する直接・間接の仕事が減る、あるいは無くなるのが考えられます。その影響についての先生の見解をお聞かせください。

- 検討

- 1 音声合成の完璧なモデルが出来たとしても、特定の声優・アーティストの音声を無断で生成して利用することはパブリシティ権侵害に該当する。
  - →したがって、結局のところ、音声合成の完璧なモデルが出来たとしても、それを利用する際には、特定の声優・アーティストの許諾が必要となり、声優等は当該許諾に際して合理的な対価を支払って貰えれば問題は無い。
  - →もっとも、力の弱い声優等はそのような対等なギャラ交渉ができず、結局「全ての権利を安く丸取りされる」というリスクがあることは否定できない。その点については、従前のプロダクション契約等でも指摘されていることであり、独禁法等で対応することになると思われる。
- 2 音声合成の完璧なモデルができ、既存のどの声優にも似ていない音声を自由に生成・利用できるようになって、当該音声に人気が出てくれば、人間の仕事が代替されることになる。
  - →この状態にどう対応するかは既存の法律の解釈からは答えが出ない。
  - →規制をかける必要があるかを新たに検討することになるが、慎重に検討する必要がある。

## 6 いくつかの具体的検討

### • 設問

- リアルな音声合成が増え実際に収録した音声との区別がつかなくなって来た時に、考えられるケースについて見解をお聞かせください。
  - ① シリーズ物のゲームで人気キャラクターにあてられていた声優・俳優の都合で降板に対して音声合成で対応するケースで起こりえそうな法的な紛争とその回避について
  - ② 収録後に不足していることが分かったセリフ、又は追加されたセリフを補うために音声合成で対応するケース
  - ③ ノンプレイヤーキャラクター(NPC)のリアルタイムで柔軟な回答を行うチャットボット的な回答に音声合成を使うケース

### • 検討

- ① 声優・俳優の許諾なく当該声優・俳優の声を音声合成で作成して利用することはパブリシティ権侵害に該当するため、対価や利用範囲についての協議・合意が必要。もっとも、今後は、降板後の事態に備えて、当初の出演契約時点で、降板した後に音声合成を利用できるか否か、及びその場合の対価についても盛り込む契約が出てくるかもしれない。
- ② ①と同じ。
- ③ 特定の声優・俳優と類似した声でなければ法的な問題はない。

## 6 いくつかの具体的検討

- 設問

- 世界中の誰でもない顔の写真を生成するAIのように、誰のものでもない音声を生成する音声合成のモデルができたとして、そのモデルの権利、生成物の権利はそれぞれどのような扱いがされられると思われますか？

- 検討

- 1 モデルの権利

- 音声合成モデルはパラメーターを組み込んだプログラムなので、当該プログラムを生成した者に当該モデルに関する権利が帰属する。

- 2 生成物の権利

- AI利用者が試行錯誤し、非常に特徴的な・魅力的な・独創的な音声を生成し、当該音声を利用して特定のセリフや歌を生成した場合には、当該セリフや歌については著作権が発生する可能性がある。一方、特定のセリフや歌を離れた、当該「音声」についての著作権は発生しないのではないか。
- また、パブリシティ権についても、人格権に由来する権利なので、AIによって生成された音声についてはパブリシティ権は発生しないと思われる。

# まとめ

- 音声合成の技術
  - 非言語音声, 存在しない話者, スタイルの合成
  - マルチトラック合成, 現在の技術の問題点
  - 今後の技術進展
- 音声合成の権利
  - 音声に関する3つの権利(セリフの著作権, 実演家の著作隣接権, 声優のパブリシティ権)
  - 設例を踏まえた検討

画像生成AIや対話AIのようなビッグウェーブが音声にもやってきます. それに備え, 技術, 倫理, 法律, 産業の専門家が連携してより良い道を模索できれば幸いです.