Vol.2022-NL-254 No.20
Vol.2022-SLP-144 No.20
2022/11/30

声を含むデータベースの「使いやすさ」に関する一考察 ～No.7音声・歌唱データベース構築を実例として～

森勢 将雅[1,a)]

概要：筆者はこれまで，音声や歌声を含む様々なデータベースを構築してきた．声を含むデータベースでは，言語に関する情報と発話音声に関する情報が混在しており，著作権に関する解釈が複雑化している．本報告では，多数存在する音声データベースから利用者が積極的に利用する目安となる「使いやすさ」について，複数の視点から議論する．議論の題材として，筆者らが構築してきたNo.7音声・歌唱データベースを用いる．

キーワード：音声・歌唱データベース，コーパス，著作権，

A study on “usability” of databases including voices As a practical example of bulding No.7 speech/singing databases

Abstract: The author has constructed several databases containing speech and singing voices. Since both linguistic information and waveform are provided in databases containing voices, copyright issues are complicated. This article discusses “usability” from several aspects, which users actively use from among the many existing voice databases. As a material for discussion, we use the No.7 speech/singing databases that we have constructed.

Keywords: Speech/singing databases, corpus, copyright,

1. はじめに

テキストから画像や音声を生成する深層学習（deep neural network; DNN）技術はすでに多数提案されており，画像・音声を生成するDNN技術は，研究者だけではなく一般ユーザにも利用されるようになった．画像生成AIではNovel AIやMidjourneyなどが2022年にリリースされ，著作権の問題を含め賛否両論の議論がなされている．AIによって生み出される創作物（以下ではAI創作物とする）については2016年に内閣官房で討議された例もあるが[*1]，急速に発展を続ける状況に対し対策が追い付いていない．

日本の著作権法第30条の4を要約すると，機械学習に利用する用途に限定すれば（つまり機械学習に利用する人間は視聴しない），著作権者の利益を不当に害しない限り著作物を利用することが可能であると解釈できる．現在はWebに多数存在するダークデータを利用する試みもあり，この著作権法はダークデータ利用と相性が良い．近年の音声認識では学習データ量が10万時間 [1] や2022年にOpenAIが発表したwhisperでは68万時間 [2] の音声が利用されているように，研究者個人で収録することが困難な大量のデータがAI研究には求められつつある．

音声認識では，学習に利用された音声の話者性などが出ることはないが，イラストや音声の生成では，学習データに用いたイラストや音声の傾向がAI創作物に反映される．特にテキスト音声合成（Text-to-speech; TTS）では学習に用いた音声の話者性が反映され，品質はすでに人間の音声と区別ができない水準に到達している [3]．前述の内閣官房の討議用資料や文献 [4] において，人工知能が自律的に生成した生成物については，現行制度上著作権の対象とは

[1] 明治大学
Meiji University, 4-21-1 Nakano, Nakano-ku, Tokyo 164–8525, Japan
[a)] mmorise@meiji.ac.jp

[*1] https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2016/jisedai_tizai/dai4/siryou2.pdf

考えられないとされている．これは，本人と区別できない音声に著作権が与えられない状況であることを意味している．かつて画像の領域で Deepfake が問題になったように，音声についても今後同様になりすまし等のリスクを考えることが必要になると考えられる．

そこで本研究では，TTS や歌声合成に利用することを前提としたデータベース（以下では DB とする）の構築を通じて，話者のリスクを下げる条件と，利用者が多数ある DB から選ぶ際に目安となる「使いやすさ」について検討してきた．ここでは，具体的な利用規約やライセンス等の複数項目について，特に TTS や歌声合成技術により話者本人に近い品質の AI 創作物が得られることも考慮した検討結果を説明する．

2. 音声・歌唱 DB の利用規約と想定される問題

音声や歌声を含む DB は複数公開されており，独自のライセンスを持つものからクリエイティブ・コモンズライセンスが設定されているものまで様々である．ここでは，いくつか代表的な音声 DB を紹介し，規約上の問題を整理する．その後，使いやすい DB の条件について説明する．

2.1 データベースのライセンス

2000 年代まで主流であった伝統的な統計的パラメトリック音声合成 [5] では，2022 年時点で最先端の TTS 技術と比較して多くの音声データを必要としていなかったため，ATR 音素バランス文 [6] のように 500 文程度のコーパスが利用されていた．なお，本稿ではテキスト情報のみを集積したものをコーパスと表記し，音声や音素ラベル，歌声では譜面データなどテキスト以外の情報を含む集積物を DB と表記する．ただし，本定義において DB と見なされる一方，提案した側がコーパスと命名しているものについては，その名称をそのまま利用する．ATR 音素バランス文は幅広く利用されてきたが，コーパスを利用することそのものに費用が発生する特徴がある．

朗読音声を Web で公開する場合，コーパスの著作権には注意する必要がある．著作権切れではない絵本の朗読動画を YouTube にアップロードすると著作権侵害になることと同様に，著作物を Web にアップロードし聴取可能な状態にすることも著作権侵害となる．あるいは，コーパスのライセンスがクリエイティブ・コモンズで，CC BY-SA のように SA が設定されている場合には注意が必要である．SA は継承を意味しており，コーパスのテキスト情報を朗読音声に変換する行為が翻案と解釈される場合[*2]，朗読音声のライセンスに影響を及ぼすリスクがある．

我々はこのような問題を回避するため，数百文からなるパブリックドメインのコーパスとして ITA コーパス [7] を構築・公開してきた．一方，新たな TTS 技術として WaveNet [8] が提案された 2016 年以降は，自然な音声を生成するために必要な音声量が増加している．そのため音声 DB も収録量の大規模化が進められており，日本語の音声 DB では JSUT コーパス [9] が 10 時間程度の大規模なものとして公開されている．JSUT コーパスは，読み上げる文リストの一部のライセンスに CC BY-SA が含まれるため，朗読音声を公開する場合のライセンスへの影響が未知数という点が懸念となる．

2.2 本人と等価な話者性を持つ音声合成への対策

最先端の TTS 技術はすでに本人の発話のコピーと言える AI 創作物を出力でき，これは TTS ソフトウェアを用いることで話者のなりすましが可能であることを意味する．加えて，AI 創作物に著作権が認められない現状では，本人と区別ができない音質で著作権の無い音声を無制限に公開することが可能である．

この問題は，TTS ソフトウェアの利用規約として禁止することが望ましいが，ソフトウェア開発者が利用規約を厳密に整備するとは限らない．対処するためには，音声 DB の利用規約の段階で，音声 DB を利用した TTS ソフトウェアに対し特定の利用規約を設けるよう，音声 DB の提供側が間接的な規約まで設定することが求められる．

2.3 使いやすさに対する目的設定

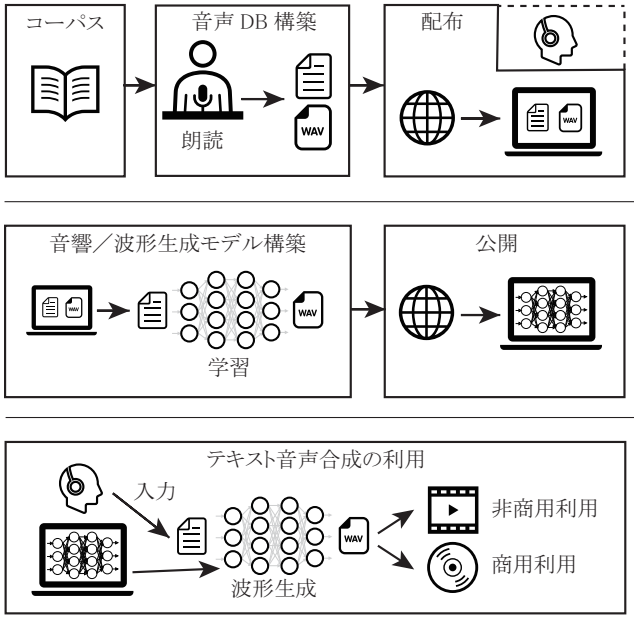
音声 DB を組み込んだ TTS や歌声合成ソフトウェアの利用範囲に制約をかけることは，声の肖像権[*3]を守ることに繋がる一方，厳しすぎる利用範囲の設定は使いやすさを損なう．とりわけ，現在の TTS や歌声合成ソフトウェアが利用される主な場はニコニコ動画や YouTube でのコンテンツ利用であることから，これらでの利用を禁止，あるいは有償利用とする条件は，ソフトウェアを使ってほしい開発者にとっては敬遠する理由になりうる．AI 生成物が本人と等価な品質であることは，不適切な利用により社会的信用を損なうリスクに繋がる．話者がプロ声優である場合，本人と等価な AI 生成物を無償で完全に制限無く利用されることは，自身の仕事と競合し利益を損なうと判断されるリスクもある．

話者，ソフトウェア開発者，ソフトウェア利用者全ての希望を完璧に満たすことは難しいため，音声 DB 構築時点で全体のバランスを考慮した利用規約を設定することが重要となる．このバランスを考え実践した結果が，筆者らが構築した「No.7」というキャラクターの音声・歌唱 DB[*4]である．以下では，No.7 の音声・歌唱 DB に対象を絞り，上記のバランスをどのように調整したのかを説明する．

[*2] この解釈が妥当かどうかは現時点で不明である．

[*3] 音響学会の音のなんでもコーナー Q74 に，声に著作権や肖像権が認められるということが記載されている．

[*4] https://voiceseven.com/



音声 DB 構築時の注意点

1. 音声 DB に影響しないコーパスの著作権とライセンス
2. 収録音声の著作権保持者
3. Web 配布時にトラブル対応が可能な音声配布の条件

音声 DB 利用者に関する制約

4. 機械学習の用途以外に利用することを許可する場合，利用を許可する範囲（著作権法第 30 条の 4 を考慮）
5. 構築した音響／波形生成モデルを含むソフトウェア公開の可否（商用・非商用それぞれに設定）

ソフトウェア利用者に関する制約

6. ソフトウェアにより生成された音声に対する商用・非商用利用それぞれの規約
7. 音声提供者者の肖像権を守ることを目的とした発話内容に関する規約

図 1 音声 DB 構築において注意した主な項目

3. DB 構築時に整備した規約

ここからは，No.7 音声・歌唱 DB において注意した項目を順番に述べる．図 1 が音声 DB 構築に限定して利用時にどのようにデータが流れていくかを示す全体像である．

3.1 音声・歌唱 DB の概要

2021 年 7 月 7 日に公開した歌唱 DB は NEUTRINO[*5]に組み込まれ，2022 年 9 月 30 日に公開した音声 DB は VOICEVOX[*6]に組み込まれている．歌唱 DB 構築時からコンセプトとして「使いやすさ」を重視しており，その一環として，知名度が高い声優を起用することに加え，声だけではなく初音ミクのようにキャラクター画像を設定することとした．両 DB の概要は表 1 に示すとおりである．

音声 DB では，モーラバランスを調整した 4,600 文からなるパブリックドメインの ROHAN コーパス [10] を採用し，3 つの発話スタイルで収録している．収録時のリップノイズ等の除去もサウンドエンジニアにより行われているが，雑音除去をしていない生データも併せて公開している．配布しているファイル数は，4,600 文× 3 スタイルに雑音除去の有無 2 種類を乗じた 27,600 文となる．データサイズが大きいため，量子化ビット数を 24 bit に調整したものを配布対象としている．また，TTS への利用に向けてフルコンテキストラベルを配布することが望ましいが，ROHAN コーパスではモーラバランス調整のため通常の日本語に出現しない海外の地名などが頻出する．これは，アクセント情報を辞書等に基づいて与えることが困難であることを意味するため，本 DB では HTS により自動で音素境界情報を与えたモノフォンラベルのみ配布することとした．

表 1 No.7 音声・歌唱 DB の概要

共通条件	
話者	小岩井ことり
収録環境	レコーディングスタジオ
マイクロフォン	NEUMANN U 87 Ai
サンプリング	96 kHz/32 bit
音声 DB の条件	
発話スタイル	ノーマル，アナウンス，絵本の読み聞かせ
収録音声数	13,800 文（各スタイルにつき 4,600 文）
コーパス	パブリックドメインの ROHAN コーパス [10]
前処理	雑音除去
その他配布物	モノフォンラベル
補足	配布時は量子化ビットを 24 bit に変更
歌唱 DB の条件	
収録楽曲	話者が独自に作曲した 51 曲
収録時間	合計 60 分程度
前処理	雑音除去，ピッチ／タイミング補正
その他配布物	musicXML と MIDI とモノフォンラベル
補足	楽曲はメロディラインのみ

歌唱 DB では，話者自身が作曲することで著作権の問題をクリアした 51 曲[*7]を対象に，話者自身の歌声を収録した．楽曲の著作権は DB 開発側に譲渡せず，作曲者自身が持ち続ける条件にしている．歌詞やメロディラインのエントロピーをある程度重視することとし，楽曲は 1 曲あたり 1 分強で合計 60 分程度になるよう調整した．これは，既存の歌声合成研究において 30 分程度の歌声から自然な歌声が得られているという知見 [11] をベースに，東北きりたん歌唱 DB [12] と同程度の規模を目指した結果である．No.7 歌唱 DB の詳細については，論文 [13] で説明されている．

*5 https://studio-neutrino.com/
*6 https://voicevox.hiroshiba.jp/
*7 1 曲は高音域のロングトーンをまとめたものである．

3.2 音声 DB 構築時の注意点

以下では，主に音声・歌唱 DB 両方に共通する構築時の注意点をまとめる．図 1 に示しているように，TTS・歌声合成への利用を想定していることから，音声 DB に関する規約に加え，TTS ソフトウェアとして配布する条件やソフトウェアにより生成された音声・歌声の利用範囲までカバーすることにした．なお，以下には著作権等に関する内容が含まれるが，確定した内容ではないことに注意する必要がある．

3.2.1 コーパスのライセンス

音声 DB 構築において最初の注意点は，朗読対象となるコーパスのライセンスである．音声 DB ではコーパスの読み上げではなく自由会話を収録したものも存在するが [14]，TTS 利用では安定した声質での発話が要求されるため [15]，TTS 用の音声 DB では現時点でコーパス朗読が妥当である．音声 DB で朗読するコーパスについては，CC BY-SA の設定で問題が必ず生じるとは言えない一方，リスクが 0 ということも同様に言えない．

確実に安全なコーパスのライセンスはパブリックドメイン（CC 0）となるため，No.7 音声 DB では，上記の問題を勘案して構築したパブリックドメインの ROHAN コーパス [10] を用いることで，この問題を回避している．ROHAN コーパスは文章量も 4,600 文と多く，モーラの出現頻度の調整もされているため，日本語では出現しないモーラまで一定量カバーできるという利点がある．

3.2.2 収録音声の著作権

音声 DB を組み込んだ TTS ソフトウェアを開発するなどの商用利用を考える場合，収録音声の著作権を DB 開発者側に譲渡しているかは重要である．著作権を譲渡していない場合は，著作権の利用について話者，あるいは話者が事務所所属のプロ声優の場合は，事務所を通した交渉が必要になる．

ソフトウェア開発に向けた交渉の手間を削減するため，No.7 音声 DB では収録音声の著作権は DB 開発側に譲渡し，音声 DB の利用範囲については 3.3.2 項で述べる規約で定める形とした．具体的に，非商用であればソフトウェアの配布を認め，商用利用の場合でも，禁止するのではなく音声 DB の転用料を支払うことで認めることとした．

3.2.3 音声 DB の配布方法

TTS ソフトウェアに組み込まれることを前提として音声 DB を配布する場合，AI 創作物の不適切な利用によるリスクを話者が負うことになる．このリスクは利用規約を厳しくすることでも対応できるが，規約違反者に対して速やかに対処できるようにすることもリスク低減に繋がる．

No.7 音声・歌唱 DB では，ダウンロード前に Twitter での認証を求めているが，これは利用規約に同意することを担保することに加え，万が一トラブルが生じた際にダウンロード者をトラッキングする狙いがある．Twitter では偽名やダミーアカウントでの登録も可能ではあるが，かといって実名により誓約書を提出させるような運用では使いやすさが損なわれる．そのため，万全ではないもののある程度のリスク管理をし，トラブル発生時には DB 開発者側もトラブル対処に動くことを条件に，音声・歌唱 DB 構築の許可を得ることとした．

3.3 音声 DB 利用者に関する制約

音声 DB の利用目的が音声認識であれば，機械学習の結果構築される音響モデルから話者を特定できる生成物が出力されることがない．一方 TTS の場合は話者性に意味があり，AI 創作物には学習に用いた話者性が表れる．そのため，TTS で利用されることを前提とする場合は，利用される範囲について注意深く設定する必要がある．

3.3.1 音声 DB の利用範囲

今回構築した No.7 音声 DB では問題にならないが，著作権法第 30 条の 4 に基づく利用を考える場合は利用範囲の設定に注意が必要である．具体的には，公開時において用途を「著作物に表現された思想又は感情の享受を目的としない利用」と陽に記載し，送信可能化権への対応として生データではなく圧縮したファイルを公開する形となる．このやり方は，過去に構築した東北きりたん歌唱 DB [12] で実施しており，現時点で問題は報告されていない．

コーパスや歌詞の利用で問題が生じない No.7 音声・歌唱 DB では，音声も歌声も聴取することを前提とした利用も認められている．例えば，きりたん歌唱 DB では認められない例として，歌声合成の結果を収録した歌声と聴取実験により比較することが挙げられる．研究用途であれば TTS 以外でも自由に使えるという条件は，使いやすさを向上させるために重要である．

3.3.2 TTS ソフトウェア配布の可否

2020 年代に入り提案されている様々な TTS 技術を用いることで，学習に十分な音声を用いるという条件はあるものの人間と等価な品質での音声生成が可能になっている．ここで懸念すべき点は，話者のなりすましや名誉棄損となるような発言を生成できてしまうことである．一番厳しい対策は，音声 DB の利用規約に，機械学習により得られた生成モデルの配布を禁止するという規約を含めることである．この規約は，研究者が研究用途で用いる場合であれば問題ないと言える一方，製品化や開発した TTS ソフトウェアを一般ユーザに使ってほしい開発者にとっては，利用するモチベーションに悪影響を与える．

使いやすさを考える場合，可能な限り多くのユーザが制約無く利用できる状況が望ましい．ソフトウェアユーザ視点では，配布した TTS ソフトウェアを用いて生成した様々な音声を公開できることは認めるべきであると言える．他方，プロ声優等が音声を提供する場合，音声 DB を用いた商品化が無償でできてしまうことは，自身の業務と競合す

る可能性がある．音声研究に対し音声を提供することそのものが声優業界としてリスクと判断されることは，音声分野の発展を考えると望ましいとは言い難い．

折衷案として，No.7 音声・歌唱 DB では，ソフトウェアにより生成した音声に利用規約を定める形で商用化を目指さないフリーソフトでの配布を認めることとした．TTS ソフトウェアの商用化を目指す場合は事前に商用利用を希望する旨連絡してもらい，必要に応じて有償で対応することとした．

3.4 ソフトウェア利用者に関する制約

音声 DB を利用した TTS ソフトウェアの配布を認める場合，ソフトウェアにより生成された様々な音声が多くの人に聴取されることも同時に認めることになる．ソフトウェアにより生成された音声の著作権は今のところ認められず [4]，AI 創作物によりどのような問題が生じるかは未知数といえる．したがって，音声 DB 構築時の段階から，TTS ソフトウェアで生成された音声に対する利用規約を設定することが現実的な対策となる．

3.4.1 ソフトウェアにより生成された音声の利用

前述のとおり，No.7 音声 DB は VOICEVOX に，歌唱 DB は NEUTRINO に利用されており，それぞれのソフトウェアで生成された音声・歌声について

(1) 利用する作品の著作者の社会的な評価を損なうような利用
(2) 他者の権利を侵害する，または侵害のおそれがある利用

等を禁止事項として定めている[*8]．必要に応じて利用許諾を中止する規約を含めることは，話者を守るために必要不可欠である．

現状では，YouTube やニコニコ動画に投稿すると広告収入が入る可能性もあり，歌声合成ソフトウェアによる AI 創作物を用いた音楽を同人イベントで頒布できることも使いやすいソフトウェアの条件に含まれる．そこで，個人利用に限定するが，同人イベントでの頒布や広告収入を得ることについて，製作にかかったコスト程度の儲けを原価回収として商用範囲と見なさない，という特殊なルールを設定した．このルール設計については，東北ずん子プロジェクトにおけるキャラクター利用の手引き[*9]を参考にしている．

広報展開を考えると，商用利用を一律で禁止することはソフトウェアユーザに対する使いやすさに悪影響を及ぼす．そのため，AI 創作物を商用利用する場合は，音声 DB の商用利用と同様に事前に連絡をし，必要に応じて利用契約を結ぶことを条件としている．この商用利用は，テレビやラジオでの利用，企業の製品説明プレゼン，ゲームの台詞やナレーション等を想定している．

[*8] 詳細は https://voiceseven.com/#j0400 に記述がある
[*9] https://zunko.jp/guideline.html

3.4.2 商用利用に関して

声の肖像権を認められると解釈した場合でも，計算機により生成された音声に対して当てはまるかには議論の余地がある．加えて，仮に TTS ソフトウェアの生成結果が本人の音声と見なされる状況になると，著作隣接権に影響する可能性がある．これは，TTS ソフトウェアの生成結果が，実演家の権利を侵害するかという論点でなされる．どちらの場合においても，話者にリスクについて事前に説明することは必須であり，その上で許容されるリスクと納得して契約してもらうことになる．

TTS ソフトウェアを配布することは，DB 開発者だけではなく DB 利用者によりなされることも想定される．それら TTS ソフトウェアで生成された音声や歌声の商用利用に費用が発生することは，前述の話者の実演家の権利に基づく．AI 創作物が話者のコピーであると解釈すると，TTS ソフトウェアによる出力は話者による実演をコピーしたという解釈も考えられる．そのため，生成された音声に対する利用範囲を規約で定め，商用利用については音声 DB 開発者ではなく音声提供した話者（あるいは所属する事務所）と契約を結ぶことが妥当であると判断した．このような契約の形にすることは，利用側の利便性がやや損なわれる一方，音声を提供する側にも利益が生じるというメリットがある．

4. データベースの「使いやすさ」に関する議論

ここでは，No.7 音声・歌唱 DB を対象に，既存の DB に対する使いやすさについて，いくつかの視点から議論する．

4.1 研究開発者にとっての使いやすさ

音声 DB を用いて TTS ソフトウェアを開発する場合，フリーソフトであればソフトウェアを自由に配布できる条件は，ソフトウェア利用者にとって使いやすいが製品開発者にとってはやや使いにくい規約と考えられる．ただし，研究用途としてであれば自由に使えるため，将来的に音声 DB を用いて製品開発をする場合においても，製品化の可能性を先に探ることが可能である．検証後，製品化の見込みがあると判断した後で商用利用の申請を行えることは，開発者としても新規で収録する必要が無いため，ある程度は使いやすい音声 DB だと考えている．

日本語の歌声合成については，著作権法の改定が 2019 年でありそれまでは著作権切れではない歌声の公開はできなかったため，公開されている歌唱 DB の数は音声 DB と比較して少ない．No.7 歌唱 DB は，著作権の問題もなく研究用途であれば自由に使えるため，歌唱合成研究のベンチマーク用の基盤として利用することが可能である．東北きりたん歌唱 DB とは異なり著作権法第 30 条の 4 に基づく用途に限定せず利用できることは，本歌唱 DB の強みであると言える．

4.2 ソフトウェア利用者にとっての使いやすさ

ソフトウェアにより生成された音声・歌声について，節度を守った同人利用を非商用と見なし事前申請無く利用できる規約は，利用者にとって比較的使いやすい考えている．無制限に利用されることは声優業界のダンピングに繋がるという懸念があることから，主に企業を対象とした営利目的の利用を商用利用とし費用を請求できる規約を設けることで，音声 DB の構築が声優の利益を損なわないよう配慮している．

DB 利用の段階から TTS ソフトウェアにより生成された AI 創作物の規約まで見通すことは，今後声優などプロを雇用した音声 DB を構築する際には重要であると考えている．ソフトウェア利用者にとってはやや利便性が損なわれることになるが，音声提供側の利益も確保する契約を結ぶことは，音声研究者と声優業界とが共存する戦略としての意味がある．

4.3 今後注意すべき点

画像生成 AI や Deepfake 等，画像に関する分野ではたびたび SNS での炎上も起きており，音声についても AI 美空ひばりや安倍晋三元総理追悼 AI プロジェクトのように何度か議論になっている．音声 DB を構築する側は，音声提供者の利益や名誉を損なうことがないように注意することが重要であり，No.7 音声・歌唱 DB の運用を話者の許可を得て実施することは，1 つの社会実験と位置付けている．

現時点でトラブルの報告は上がってきていない一方，ニコニコ動画や YouTube への動画投稿も定期的にあり商用利用に向けた問い合わせも届いていることから，規約が有効に機能していると思われる．将来トラブルが起きた際にこの契約のスタイルで問題が無いか実例を示すことは，音声生成 AI の未来にも資すると考えている．

5. おわりに

本稿では，音声や歌声を含む DB 構築に対し，収録時間等の量や音素バランス等の質ではなく，著作権や利用用途等の規約を対象とした「使いやすさ」の観点から考察を述べた．本研究で構築してきた音声・歌唱 DB の利用において，現時点でトラブルは報告されていないが，今後問題が生じる可能性もある．その際には，本稿で述べた考察の妥当性等が議論されることになり，結果として DB 構築における注意点の事例が蓄積されるだろう．

TTS の場合，ある程度の SNR が確保され発話時の話者性や感情が統一されていることが望ましいとされており，その意味で音声 DB には現在も十分な価値がある．一方，技術革新により音声のダークデータが TTS でも利用可能になれば，将来的には音声合成用の DB やコーパスは，存在意義を失う可能性もある．その際にも話者の肖像権等を含む問題が生じる可能性はあることから，引き続き様々な使いやすい音声 DB を公開することで，どのような問題が生じるかの検証を続ける予定である．

謝辞 本研究の一部は，科研費 JP21K19794, JP21H04900 の支援を受けた．

参考文献

[1] H. Soltau, H. Liao, and H. Sak, “Neural speech recognizer: Acoustic-to-Word LSTM model for large vocabulary speech recognition,” in Proc. INTERSPEECH2017, pp. 3707–3711, 2017.
[2] A. Radford, J.W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, “Robust Speech Recognition via Large-Scale Weak Supervision,” `https://github.com/openai/whisper`.
[3] J. Shen, R. Pang, R. J. Weiss, M. Schuster, N. Jaitly, Z. Yang, Z. Chen, Y. Zhang, Y. Wang, RJ Skerry-Ryan, R. A. Saurous, Y. Agiomyrgiannakis, and Y. Wu, “Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions,” in Proc. ICASSP 2018, pp. 4779–4783, 2018.
[4] 愛知靖之，“AI 生成物・機械学習と著作権法，” パテント，vol. 73, no. 8, pp. 131–146, 2020.
[5] H. Zen, K. Tokuda, and A. W. Black, “Statistical parametric speech synthesis,” Speech Communication, vol. 51, no. 11, pp. 1039–1064, 2009.
[6] 小林哲則，板橋秀一，速水悟，竹澤寿幸，“日本音響学会研究用連続音声データベース，” 日本音響学会誌，vol. 48, no. 12, pp. 888–893, 1992.
[7] 小口純矢，金井郁也，小田恭央，齊藤剛史，森勢将雅，“ITA コーパス：パブリックドメインの音素バランス文からなる日本語テキストコーパスの構築と基礎評価，” 情報処理学会音楽情報科学研究会, vol. 2021-MUS-131, no. 31, pp. 1–4, 2021.
[8] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “WaveNet: A generative model for raw audio,” arXiv preprint arXiv:1609.03499, 2016.
[9] S. Takamichi, R. Sonobe, K. Mitsui, Y. Saito, T. Koriyama, N. Tanji, and H. Saruwatari, “JSUT and JVS: Free Japanese voice corpora for accelerating speech synthesis research,” Acoust. Sci. Tech, vol. 41, no. 5, pp. 34–45, 2021.
[10] 森勢将雅，“ROHAN：テキスト音声合成に向けたモーラバランス型日本語コーパス，” 日本音響学会誌，vol. 79, no. 1（2023 年掲載予定）.
[11] M. Blaauw and J. Bonada, “A neural parametric singing synthesizer,” in Proc. INTERSPEECH 2017, pp. 4001–4005, 2017
[12] I. Ogawa and M. Morise, “Tohoku Kiritan singing database: A singing database for statistical parametric singing synthesis using Japanese pop songs,” Acoustical Science and Technology, vol. 42, no. 3, pp. 140-145, 2021.
[13] 森勢将雅，藤本健，小岩井ことり，“レアなモーラを含む日本語歌唱データベースの構築と基礎評価,” 情報処理学会論文誌, vol. 63, no. 9, pp. 1523-1531, 2022.
[14] H. Mori, H. Kasuya, M. Nakamura, and M. Amanuma, “Some considerations for designing spoken dialogue database from the viewpoint of paralinguistic information,” Acoust. Sci. Tech, vol. 24, no. 6, pp. 376–378, 2003.
[15] 山本龍一，高道慎之介，“Python で学ぶ音声合成，” 株式会社インプレス，2021.