

令和6年度特許出願技術動向調査

—メタバース時代に向けた音声・音楽処理—

Patent Trends in Speech and Audio Processing



特許庁 総務部企画調査課 知財動向班

中村 俊之

平成26年4月特許庁入庁（審査第三部）。環境化学、無機化学、生命工学にて審査業務、情報技術統括室、福島ロボットテストフィールド、調査室での業務、米国コネチカット大学への留学等の経験を経て、令和7年7月より現職。

1 特許出願技術動向調査について

特許出願は公報として広く一般に公開される。特許の公開情報は、企業・大学等における研究開発の成果に係る技術情報や権利情報である。これらの特許情報から、先端技術分野等の特許出願状況や研究開発の方向性を分析することで、企業、大学等における今後の研究開発の方向性を決定する指針となる。そこで、特許庁では、国の政策として推進すべき技術分野、社会的に注目されている技術分野等から技術テーマを選定して、平成11年度から「特許出願技術動向調査」と題し特許出願の分析を行っている。

令和6年度は、5の技術テーマを選定し調査を行っており、本稿ではその中から、近年注目度の高い「メタバース時代に向けた音声・音楽処理」についての調査の概要を述べる。

2 本調査の調査結果

2.1 概要

音を記録する技術や音を再現する技術は現代社会において重要であり、かつ人々の生活にも密接に関わっている。音に関する製品やサービスの態様も、近年大きく変化しており、例えば、音楽を聴くために以前はCDを利用していたが、現在ではインターネットを利用した配信サービスが主流になっている。また、AI技術の発展に支えられた音声合成や自動作曲に基づくサービスが次々に登場し、一般にも利用できるようになった。

こうした技術の発展は、音声や音楽が活用される場面を多様化させることにつながる。その一例としてメタバースが挙げられる。一方、技術の発展に伴い、新たな課題も生じている。AIによって著名人の声が模倣され悪用されるといった問題や、AIが他人の著作物に似た楽曲を生成するといった著作権に関する問題が挙げられる。また、AIそのものに対するリスクも認識され、その対応が議論されている。

音声や音楽を取り扱う際、このような課題にも向き合いつつ、人々にとって有用な音声や音楽を届けるための技術を提供していくことが必要である。そこで、本調査では、音声・音楽処理技術に関する研究開発、技術開発の状況を把握し、将来に向けてどのような方向性を定めることが必要なのかを検討するために、特許出願や論文発表の動向を調査した。



図1 技術俯瞰図（図1-1-1）

図1に本調査における調査対象範囲を表す技術俯瞰図を示す。この技術俯瞰図は、音がどのように生成され、出力され、人間の耳に届くのか、という流れに着目し、

各段階においてどのような技術要素があるのかを説明したものである。

① 入力

音声または音楽を生成するためには入力情報が必要である。入力情報としては、音声、楽器音、文章（テキスト）などが挙げられる。そのほか、画像や映像をもとに自動で音楽を生成する技術も開発されており、この場合、画像や映像の情景、雰囲気などを入力情報とすることで、その画像や映像にふさわしい音楽を生成することができる。また、脳波を読み取り、これを入力情報として合成音声を生産する研究も見られる。

② 生成処理

入力情報に基づいて音声または音楽の生成処理が行われる。具体的には、音声変換や歌声変換、音声合成や歌声合成などの音声処理が行われる。本調査では、入力情報が音声 / 歌声であって、それを処理して別の音声 / 歌声情報を生成することを「音声 / 歌声変換」と定義し、音声 / 歌声以外の入力情報から音声 / 歌声情報を生成することを「音声 / 歌声合成」と定義した。また、音楽の生成処理には、楽曲や環境音を生成するための「自動作曲」技術、「作曲支援」技術が含まれる。これらの生成処理には、入力情報から学習した統計モデルを用いて音声波形を生成する統計的手法が広く利用されている。統計モデルとしては、隠れマルコフモデルやニューラルネットワークを用いたものがよく知られている。

③ 出力

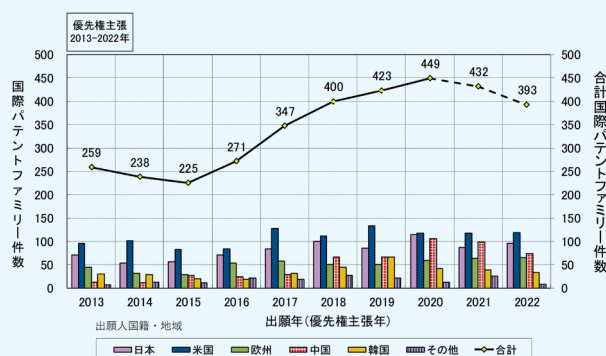
生成された音声または音楽を出力するうえで、「音場再現」技術や「音像定位」技術が重要な役割を果たす。

本調査は、上述した、音場再現・音像定位、音声合成、音声変換、歌声合成、歌声変換、自動作曲・作曲支援に関する技術を対象とした。また、本調査では、これらの技術を適用する「場面 / 領域」の側面、及び、これらの技術を利用または研究・開発するときに発生する「課題」の側面からの考察も行った。

2.2 特許出願動向

図2に出願人国籍・地域別国際 Patent ファミリー

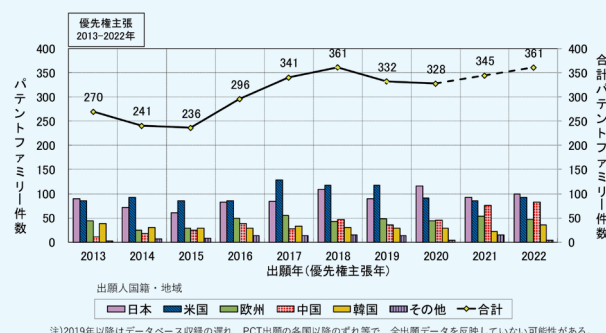
件数推移を示す。日本、米国の出願人の国際 Patent ファミリー件数が多い傾向にあるが、2018 年以降、中国の出願人の国際 Patent ファミリー件数増加が確認できる。これ以降、本稿では技術区分「音場再現・音像定位」、「音声変換」、「音声合成」及び「社会的課題」に焦点を当てて紹介する。



注) 2019 年以降はデータベース収録の遅れ、PCT 出願の各国以降のずれ等で、全出願データを反映していない可能性がある。

図2 出願人国籍・地域別国際 Patent ファミリー件数推移 (図 4-2-2)

図3に音場再現・音像定位における出願人国籍・地域別 Patent ファミリー件数推移を示す。他の技術区分に見られるような大きな増加は見られず、定常的に、技術開発が一定の特許出願に結び付いている様子が見られる。



注) 2019 年以降はデータベース収録の遅れ、PCT 出願の各国以降のずれ等で、全出願データを反映していない可能性がある。

図3 出願人国籍・地域別 Patent ファミリー件数推移 (「音場再現・音像定位」) (図 4-3-1)

図4に音声変換における出願人国籍・地域別 Patent ファミリー件数推移を、図5に音声合成における出願人国籍・地域別 Patent ファミリー件数推移をそれぞれ示す。音声変換、音声合成ともに、2018 年以降の Patent ファミリー件数増加が顕著である。2016 年に発表された WaveNet の論文が大きく影響を与えているものと考えられる。WaveNet は音声波形を直接生成するディープニューラルネットワークであり、より自然な発音に近い音声を合成することに成功したニューラ

ルネットワークモデルである。これを利用した技術開発の成果が特許出願に結びついていると考えられる。

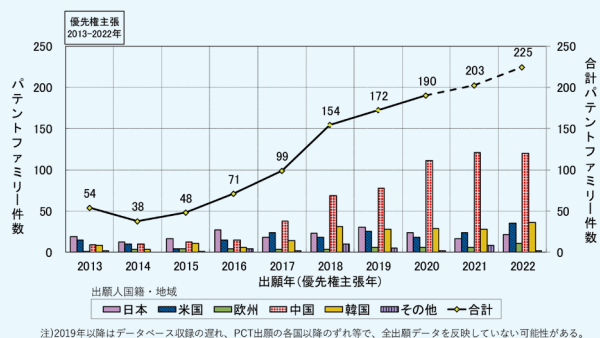


図4 出願人国籍・地域別特許ファミリー件数推移（「音声変換」）（図4-4-1）

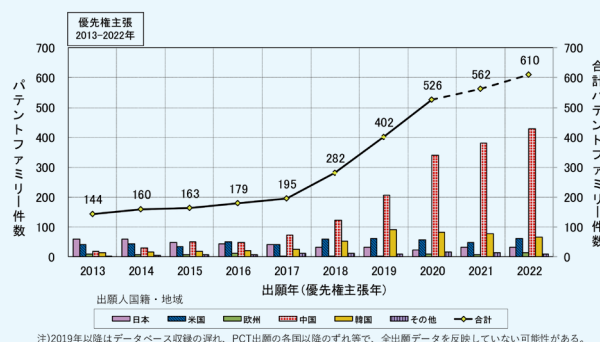


図5 出願人国籍・地域別特許ファミリー件数推移（「音声合成」）（図4-5-1）

図6に技術区分「社会的課題」別特許ファミリー件数推移を示す。「権利保護」に関する特許ファミリー件数は他の区分に比べると少ないものとなっていることが分かる。

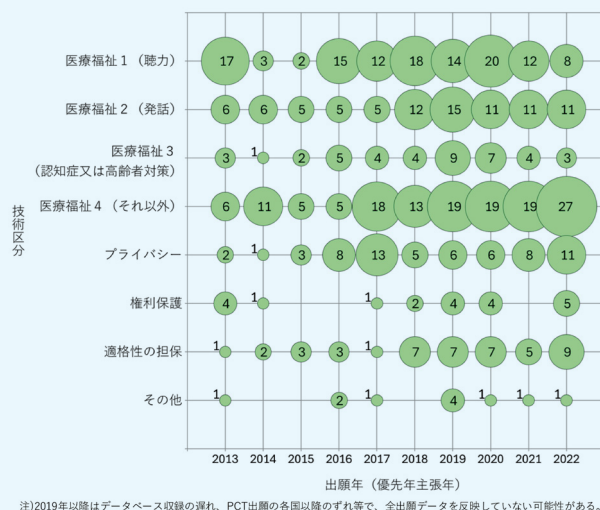


図6 技術区分「社会的課題」別特許ファミリー件数推移（図4-9-1）

2.3 クロス集計による特許技術分析

図7に本調査の技術区分の大区分である音場再現・音像定位、音声変換、音声合成、歌声変換、歌声合成、自動作曲・作曲支援の6つの技術区分のうち、メタバースを意識した（メタバースに関する技術区分にも分類された）特許ファミリーがどの程度存在するかを表すグラフを示す。メタバースを意識した特許出願は音場再現・音像定位の技術区分において多く出願されていることが分かる。その他の技術区分ではこれに比べると少数であり、メタバースを意識した技術開発活動は盛んではないものと考えられる。

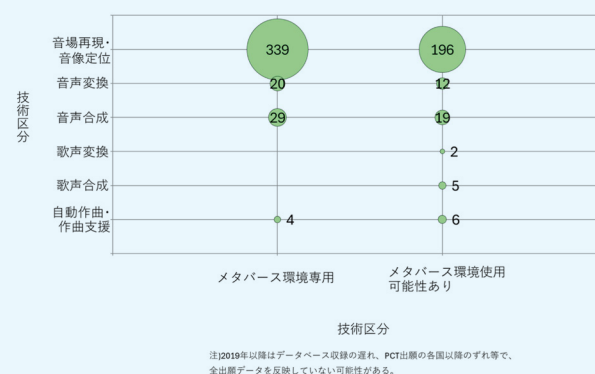


図7 技術区分「大区分」別技術区分「メタバース」別特許ファミリー件数（図4-10-1）

2.4 研究開発動向

図8に研究者所属機関国籍・地域別の論文発表件数推移を示す。2021年以降は中国からの論文発表が大きく増加し、年当たりの発表件数は欧州と同程度となっている。

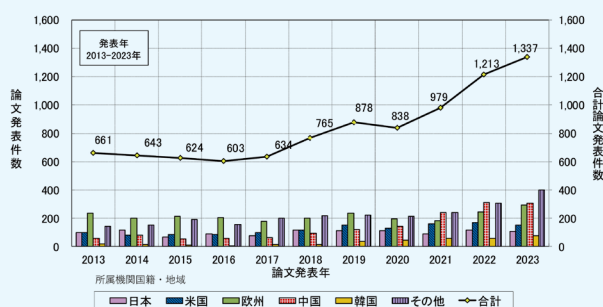


図8 研究者所属機関国籍・地域別論文発表件数推移（図5-2-1）

2.5 まとめ

以上の調査結果を踏まえ、日本の目指すべき方向性に関して検討を行い、その検討結果を、明確な方向性を示す「提言」、及び、今後に向けて留意する必要があると思われる傾向を示す「示唆」の形でまとめた。

2.5.1 提言1：音の伝達に付加価値を与えるための研究開発、技術開発を継続し、日本が今後も存在感、影響力を高めていくことが必要

音声・音楽処理技術においては、音をどのように聴取者へ伝達するのか、という視点が重要である。その基礎的な技術として、音像定位技術、音場再現技術がある。特許出願動向を見ると、音像定位技術、音場再現技術ともに毎年一定の量の特許出願が確認できる。また、研究開発動向を見ると、論文発表件数も毎年一定量が確認できる。

これらのことは音の伝達に対して付加価値を与えるための研究開発、技術開発活動が常に行われていることを示している。音場再現・音像定位技術に関する出願人国籍・地域別パテントファミリー件数は、日本からの出願が多くなっていることから、これらの技術において日本は既に一定の存在感、影響力を持っているものと推測される。今後、音の伝達の場面はメタバースを始めとして広がっていくと考えられることから、この点の研究開発、技術開発活動を今後も継続し、日本の存在感、影響力を高めていくべきである。

2.5.2 提言2：音声合成技術の実用化に際して、社会情勢やそれに基づいた政策意思決定の動向を確認するとともに、悪用防止の研究開発、技術開発へと速やかに移行できるような体制を構築することが必要

現在、社会的に影響力のある人物の声を使い、その人物になりすまして電話をかけるといった事案の発生が確認されている。このようなことは詐欺等の好ましくない行為に利用される恐れがあるとの指摘がなされている。しかし、本調査の結果、権利保護という社会的課題に関連した特許出願は少なく、なりすましやディープフェイクを技術的に防ぐための技術開発活動は盛んではないことが分かる。この背景として、これらの技術開発活動が企業の利益に直結するわけではないため、企業が積極的に投資しづらいという可能性や、公益的観点から企業が権利取得を見送るという可能性が考えられる。

一方、声のなりすましを許さないという社会的な機運が高まっている。これを背景としてか、権利保護に関する論文の発表件数は2021年以降に増加傾向を見せており、今後の研究開発活動の活発化が期待される。欧州では、2024年5月にEU理事会によってAI法案が採択され、音声コンテンツを生成または操作するためにAIシステム

を利用する場合、そのコンテンツがAIによるものであることを開示する必要があることが規定された。このように社会的な機運によって政策面での指針が策定されており、これに対応する研究開発活動、技術開発活動が進展していくと考えられる。日本においても、社会情勢やそれに基づいた政策意思決定の動向を注視し、これらに対応した研究開発活動、技術開発活動の開始に乗り遅れないための準備をしておくことが望ましいといえる。

2.5.3 提言3：音楽知識がなくとも作曲の喜びを提供する研究開発、技術開発を進めることによって、今後の市場において優位性を確保することが必要

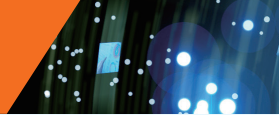
音楽は誰でも楽しむことができるエンターテインメントであるが、それを自ら作成することや自ら楽器を演奏することには、高度な音楽知識が必要である。他方、多くの人が自ら作曲してみたい、楽器を演奏してみたいとの気持ちを持っているものと推測され、実際に、近年では複数の自動作曲サービスが市場に投入されている。

一般的に市場におけるサービスの展開においては、先行者利益が大きく、後発の参入で先行者に勝つことが難しい場合がある。2018年以降、自動作曲・作曲支援に関する中国の特許出願件数及び論文発表件数の増加が顕著であり、市場導入に向けた研究開発活動、技術開発活動が活発に行われている様子が見られる。そのような事情を鑑み、市場において他国に遅れを取らないよう研究開発活動、技術開発活動を進める必要があるといえる。

2.5.4 示唆1：

音声処理技術、音楽処理技術の進展には技術革新が非常に大きく関わっていることが示唆された。革新的な技術あるいは理論の登場は、流通する製品あるいはサービスを大きく変化させる可能性が高いため、そのような技術や理論をいち早く製品、サービス開発に取り入れることによって、他に先駆けて消費者に支持される製品、サービスを提供することが必要である。

音声変換、音声合成における特許出願技術動向調査、研究開発動向調査の結果では、2016年以降に大きく特許出願件数及び論文発表件数が伸びていることが示された。2016年はWaveNetに関する論文が発表された年であり、これを機に研究開発活動、技術開発活動の風向きが変わったものと考えられる。こうした理論の発表後



に特許出願、論文発表が増加することは、最新の研究開発、技術開発動向に気を配り、その成果をいち早く製品、サービス開発に取り入れることが産業競争力の確保に必要であることを示している。したがって、旧来から存在する技術や理論に対するブレイクスルーの登場に注意し、そのような技術や理論を迅速に取り入れて研究開発活動、技術開発活動を進めて行くことが望ましいといえる。

2.5.5 示唆2：

音声処理技術、音楽処理技術の展開先として、メタバースが挙げられる。メタバースの基盤となるプラットフォームの構築においては、現在のインターネットの利用状況を見る限り、米国の影響力が強いと考えられる。日本は音場再現・音像定位においては米国に次ぐパテントファミリー件数があり、この分野において一定の存在感を有しているといえる。メタバースへの展開を見据えた場合、これらの技術を活用し、産業競争力を確保していくことが考えられる。

音場再現・音像定位、音声変換、音声合成、歌声変換、歌声合成、自動作曲・作曲支援の各技術について、メタバースを意識した出願がどの程度あるのかを調査したところ、音場再現・音像定位においては他の技術区分に比べてメタバースを意識する傾向が強く表れていた。メタバースでは臨場感や没入感が重要であると考えられていることから、音を伝達するための基礎的な技術ではメタバースも意識した技術開発が行われているものと考えられる。日本は音場再現・音像定位における特許出願に強みを持っていることから、メタバースへの展開に関しても、この音場再現・音像定位を基礎として研究開発活動、技術開発活動を継続することが好ましいといえる。

3 結び

本稿では、令和6年度に調査を実施したテーマの中から、「メタバース時代に向けた音声・音楽処理」の調査結果を紹介した。

特許出願技術動向調査の要約版は特許庁ウェブサイトに掲載されており、報告書については国立国会図書館、特許庁図書館で閲覧可能である。

(<https://www.jpo.go.jp/resources/report/gidou-houkoku/tokkyo/index.html>)

マルチモーダル技術が知財情報にもたらす未来