

明 細 書

発明の名称：送信装置、送信方法、受信装置、及び、受信方法 技術分野

[0001] 本技術は、送信装置、送信方法、受信装置、及び、受信方法に関し、特に、視覚障害者に対するアクセシビリティを向上させることができるようにした送信装置、送信方法、受信装置、及び、受信方法に関する。

背景技術

[0002] デジタル放送の分野では、視覚障害者に対するアクセシビリティが求められている（例えば、特許文献1参照）。

[0003] 特に、米国では、いわゆる米国アクセシビリティ法（CVAA：The 21st Century Communications and Video Accessibility Act of 2010）が制定され、連邦通信委員会（FCC：Federal Communications Commission）により、この法律をベースとした映像番組のアクセシビリティに関する様々な規制が発表されている。

先行技術文献

特許文献

[0004] 特許文献1：特開2009-204711号公報

発明の概要

発明が解決しようとする課題

[0005] ところで、視覚障害者に対して、番組情報等のユーザインターフェース（UI：User Interface）を提示する場合、当該番組情報等のテキスト情報を、TTS（Text To Speech）エンジンで読み上げることで、アクセシビリティを高めるのが一般的である。

[0006] しかしながら、TTSエンジンでは、番組情報等の制作者が意図した通りに、テキスト情報が読み上げられるとは限らず、視覚障害者が、健常者と同等の情報が得られる保証がない。そのため、確実に、制作者が意図した通りの発話が行われて、視覚障害者が、健常者と同等の情報を得られるようにするた

めの技術が求められていた。

- [0007] 本技術はこのような状況に鑑みてなされたものであり、確実に、制作者が意図した通りの発話が行われるようにすることで、視覚障害者に対するアクセシビリティを向上させることができるようにするものである。

課題を解決するための手段

- [0008] 本技術の第1の側面の送信装置は、表示情報に対する制作者が意図する音声の発話に関するメタデータを生成するメタデータ生成部と、前記メタデータを含む電子番組情報を生成する電子番組情報生成部と、前記表示情報を表示可能な受信装置に対して、前記電子番組情報を送信する送信部とを備える送信装置である。
- [0009] 前記メタデータは、読み方が一意に定まらない文字列、又は発音が難解な文字列の発話に関する情報を含んでいるようにすることができる。
- [0010] 前記表示情報は、コンテンツに関する情報、又はアイコンを含んでいるようにすることができる。
- [0011] 前記コンテンツを取得するコンテンツ取得部をさらに備え、前記送信部は、前記電子番組情報を、前記コンテンツとともに、デジタル放送信号で送信するようにすることができる。
- [0012] 前記電子番組情報は、OMA-BCAST (Open Mobile Alliance - Mobile Broadcast Services Enabler Suite) で規定されたESG (Electronic Service Guide) に準拠しており、前記メタデータは、SSML (Speech Synthesis Markup Language) 形式で記述され、前記ESGを構成する所定のフラグメントに、前記SSML形式で記述された前記メタデータのファイルの取得先を示すアドレス情報、又は前記SSML形式で記述された前記メタデータの内容そのものが含まれるようにすることができる。
- [0013] 送信装置は、独立した装置であってもよいし、1つの装置を構成している内部ブロックであってもよい。
- [0014] 本技術の第1の側面の送信方法は、上述した本技術の第1の側面の送信装置に対応する送信方法である。

- [0015] 本技術の第1の側面の送信装置、及び、送信方法においては、表示情報に対する制作者が意図する音声の発話に関するメタデータが生成され、前記メタデータを含む電子番組情報が生成され、前記表示情報を表示可能な受信装置に対して、前記電子番組情報が送信される。
- [0016] 本技術の第2の側面の受信装置は、送信装置から送信されてくる、表示情報に対する制作者が意図する音声の発話に関するメタデータを含む電子番組情報を受信する受信部と、前記電子番組情報に含まれる前記メタデータを取得するメタデータ取得部と、前記メタデータに基づいて、前記表示情報を読み上げる音声読み上げ部とを備える受信装置である。
- [0017] 前記メタデータは、読み方が一意に定まらない文字列、又は発音が難解な文字列の発話に関する情報を含んでいるようにすることができる。
- [0018] 前記表示情報は、コンテンツに関する情報、又はアイコンを含んでいるようにすることができる。
- [0019] 前記受信部は、デジタル放送信号として、前記コンテンツとともに送信される前記電子番組情報を受信するようにすることができる。
- [0020] 前記電子番組情報は、OMA-BCASTで規定されたESGに準拠しており、前記メタデータは、SSML形式で記述され、前記ESGを構成する所定のフラグメントに、前記SSML形式で記述された前記メタデータのファイルの取得先を示すアドレス情報、又は前記SSML形式で記述された前記メタデータの内容そのものが含まれており、前記メタデータ取得部は、前記アドレス情報に従い前記メタデータのファイルを取得するか、又は前記フラグメントから前記メタデータを取得するようにすることができる。
- [0021] 受信装置は、独立した装置であってもよいし、1つの装置を構成している内部ブロックであってもよい。
- [0022] 本技術の第2の側面の受信方法は、上述した本技術の第2の側面の受信装置に対応する受信方法である。
- [0023] 本技術の第2の側面の受信装置、及び、受信方法においては、送信装置から送信されてくる、表示情報に対する制作者が意図する音声の発話に関する

メタデータを含む電子番組情報が受信され、前記電子番組情報に含まれる前記メタデータが取得され、前記メタデータに基づいて、前記表示情報が読み上げられる。

発明の効果

[0024] 本技術の第1の側面、及び、第2の側面によれば、視覚障害者に対するアクセシビリティを向上させることができる。

[0025] なお、ここに記載された効果は必ずしも限定されるものではなく、本開示中に記載されたいずれかの効果であってもよい。

図面の簡単な説明

[0026] [図1]番組情報やタイトルの読み上げ例を示す図である。

[図2]アイコンの読み上げ例を示す図である。

[図3]従来のTTSエンジンによるテキスト情報の読み上げ例を説明する図である。

[図4]従来のTTSエンジンによるテキスト情報の読み上げ例を説明する図である。

[図5]本技術を適用したTTSエンジンによるテキスト情報の読み上げ例を説明する図である。

[図6]本技術を適用したTTSエンジンによるテキスト情報の読み上げ例を説明する図である。

[図7]本技術を適用した放送システムの構成例を示す図である。

[図8]本技術を適用した送信装置の構成例を示す図である。

[図9]本技術を適用した受信装置の構成例を示す図である。

[図10]ESGの構造の例を示す図である。

[図11]ESGのサービスフラグメントの構成例を示す図である。

[図12]ESGのコンテンツフラグメントの構成例を示す図である。

[図13]拡張ESGの構成例を示す図である。

[図14]拡張ESGの他の構成例を示す図である。

[図15]PhoneticInfoURI要素の詳細な構成を示す図である。

[図16]PhoneticInfo要素の詳細な構成を示す図である。

[図17]SSML形式におけるsub要素の記述例を示す図である。

[図18]SSML形式におけるphoneme要素の記述例を示す図である。

[図19]SSML形式におけるaudio要素の記述例を示す図である。

[図20]送信処理を説明するフローチャートである。

[図21]受信処理を説明するフローチャートである。

[図22]コンピュータの構成例を示す図である。

発明を実施するための形態

[0027] 以下、図面を参照しながら本技術の実施の形態について説明する。なお、説明は以下の順序で行うものとする。

- [0028] 1. 本技術の音声発話メタデータの概要
2. システムの構成
3. ESGの拡張による音声発話メタデータの配置
4. 音声発話メタデータの記述例
5. 各装置で実行される処理の流れ
6. 変形例
7. コンピュータの構成

[0029] <1. 本技術の音声発話メタデータの概要>

[0030] 米国アクセシビリティ法に関係する連邦通信委員会の規制のうち、ユーザインターフェースに関する規制 (FCC Report & Order(FCC 13-138) released October 31, 2013. C.F.R. Title 47 § 79.107, 108) では、テレビ受像機等の映像番組を表示可能な受信機に対して、そのユーザインターフェースを、視覚障害者に対してアクセシブルにすることが要求されている。

[0031] 具体的には、図1に示すように、電子サービスガイド (ESG: Electronic Service Guide) の画面 (ESG画面) が表示された場合に、例えば、タイトルや番組情報などを読み上げることで、視覚障害者に対して放送番組の選局に必要な情報を提供することができる。

[0032] また、図2に示すように、メニュー画面が表示された場合には、例えば、

各種のサービスを示したアイコンについて、そのサービスの内容を読み上げることで、視覚障害者に対してサービスの内容を提供することができる。

[0033] このように、受信機に表示されるユーザインターフェースに関する情報を読み上げることで、視覚障害者は、当該ユーザインターフェースに関する情報を入手して、受信機に対する各種の操作を行うことが可能となる。

[0034] ところで、放送番組（プログラム）を選局するためのチャンネル情報やプログラム情報は、放送局などの送信機から受信機に対して、ESG情報として提供されるが、このESG情報は、主にテキスト情報やロゴデータ等からなる。そして、受信機では、ESG情報に基づいて、放送番組を選局するためのESG画面を生成して表示することになる。

[0035] 上述したように、ESG画面を表示するに際しては、そのユーザインターフェースを、視覚障害者に対してアクセシブルにすることが要求され、例えばタイトルや番組情報などを読み上げることになるが、ここでは、タイトルや番組情報等のテキスト情報を、TTS (Text To Speech) エンジンで読み上げるのが一般的である。このTTSエンジンは、テキスト情報から、人間の音声を人工的に作り出すことができる音声合成機 (Text To Speech Synthesizer) である。

[0036] しかしながら、TTSエンジンでは、ユーザインターフェースの制作者が意図した通りに、テキスト情報が読み上げられるとは限らず、視覚障害者が、健常者と同等の情報が得られる保証はない。

[0037] 具体的には、図3に示すように、例えば、“AAA”であるテキスト情報は、“triple A”又は“A A A”と読めるため、その読み方が一意に定まらないので、TTSエンジンでは、どのように読み上げてよいかを判断できず、結果として、制作者が意図した通りに、テキスト情報が読み上げられない可能性が出てくる。

[0038] また、図4に示すように、例えば、“Caius College”であるテキスト情報は、その発音が難解な固有名詞等であるため、TTSエンジンでは、どのように読み上げていいのかが判断できず、制作者が意図した通りに、テキスト情報が

読み上げられない可能性がある。

[0039] このように、テキスト情報の読み方が一意に定まらない場合や、発音が難解な固有名詞等である場合などには、制作者が意図した通りにテキスト情報が読み上げられない可能性があるため、確実に、制作者が意図した通りの発話が行われて、視覚障害者が、健常者と同等の情報を得られるようにする技術が求められていた。

[0040] そこで、本技術では、確実に、ユーザインターフェース等の表示情報に対する制作者が意図した通りの音声の発話が行われるようにするために、制作者が意図する音声の発話に関する情報（以下、「音声発話メタデータ」という。）を、TTSエンジンに提供して、当該TTSエンジンが、制作者が意図する音声を発話できるようにする。なお、当該音声発話メタデータは、ESG情報に含めて提供することができる。

[0041] 具体的には、図5に示すように、例えば、“AAA”であるテキスト情報について、その音声の読み方を示した“triple A”を、音声発話メタデータとして、TTSエンジンに提供されるようにすることで、当該TTSエンジンは、音声発話メタデータに基づいて、“triple A”と読み上げることができる。

[0042] すなわち、図3において、“AAA”であるテキスト情報を入力した場合、TTSエンジンは、“triple A”と、“A A A”のどちらで読み上げるのが正しいかを判断することができなかったが、図5においては、音声発話メタデータとしての“triple A”を入力することで、TTSエンジンは、音声発話メタデータに従い、“triple A”を読み上げることができるので、制作者が意図する音声が発話されることになる。

[0043] また、図6に示すように、例えば、“Caius College”であるテキスト情報について、その音素情報を、音声発話メタデータとして、TTSエンジンに提供されるようにすることで、当該TTSエンジンは、音声発話メタデータに基づいて、“keys college”と読み上げることができる。

[0044] すなわち、図4において、“Caius College”であるテキスト情報を入力した場合、TTSエンジンは、その発音が難解な固有名詞等であるため、どのように

読み上げるのが正しいかを判断することができなかったが、図6においては、音声発話メタデータとしての音素情報を入力することで、TTSエンジンは、音声発話メタデータに従い、“keys college”と読み上げることができるので、制作者が意図する音声が発話されることになる。

[0045] このように、音声発話メタデータをTTSエンジンに提供することで、例えば、テキスト情報の読み方が一意に定まらない場合や、発音が難解な固有名詞等である場合などであっても、確実に、制作者が意図した通りにテキスト情報が読み上げられるため、視覚障害者が、健常者と同等の情報を得られるようになる。

[0046] <2. システムの構成>

[0047] (放送システムの構成例)

図7は、本技術を適用した放送システムの構成例を示す図である。

[0048] 放送システム1は、放送番組等のコンテンツを提供するとともに、ユーザインターフェース等の表示情報を、視覚障害者に対してアクセシブルにすることが可能なシステムである。放送システム1は、送信装置10と、受信装置20から構成される。

[0049] 送信装置10は、例えば地上デジタル放送サービスを提供する放送局により運営される。送信装置10は、放送番組等のコンテンツを、デジタル放送信号により送信する。また、送信装置10は、音声発話メタデータを含むESG情報を生成して、デジタル放送信号に含めて送信する。

[0050] 受信装置20は、例えばテレビ受像機やセットトップボックス等から構成され、ユーザの各家庭等に設置される。受信装置20は、送信装置10から送信されてくるデジタル放送信号を受信し、放送番組等のコンテンツの映像や音声を出力する。

[0051] また、受信装置20は、TTSエンジンを有しており、ユーザインターフェース等の表示情報を表示する場合に、ESG情報に含まれる音声発話メタデータに基づいて、ユーザインターフェース等の表示情報を読み上げる。

[0052] ここでは、TTSエンジンが、音声発話メタデータに従い、テキスト情報等の

表示情報を読み上げることになるため、例えば、テキスト情報の読み方が一意に定まらない場合や、発音が難解な固有名詞等である場合などであっても、確実に、制作者が意図した通りに読み上げられることになる。

[0053] なお、図7の放送システム1では、1台の送信装置10のみを図示しているが、実際には、複数の放送局ごとに送信装置10が設置される。同様に、図7の放送システム1では、1台の受信装置20のみを図示しているが、実際には、複数のユーザの家庭ごとに、受信装置20が設置されている。

[0054] (送信装置の構成例)

図8は、図7の送信装置の構成例を示す図である。

[0055] 図8において、送信装置10は、コンテンツ取得部111、音声発話メタデータ生成部112、ESG情報生成部113、ストリーム生成部114、及び、送信部115から構成される。

[0056] コンテンツ取得部111は、放送番組等のコンテンツを取得して、ストリーム生成部114に供給する。また、コンテンツ取得部111は、コンテンツに対して、例えばエンコードやフォーマット形式の変換処理などを実行することができる。

[0057] なお、コンテンツとしては、例えば、既に収録されたコンテンツの保管場所から、放送時間帯に応じて該当するコンテンツが取得されたり、あるいはスタジオやロケーション場所からライブのコンテンツが取得されたりする。

[0058] 音声発話メタデータ生成部112は、例えばユーザインターフェースの制作者からの指示に従い、音声発話メタデータを生成して、ESG情報生成部113に供給する。なお、音声発話メタデータとしては、例えば、テキスト情報の読み方が一意に定まらない場合にその音声の読み方を示した情報や、発音が難解な固有名詞等である場合にその音素情報が生成される。

[0059] ここで、ESG情報に格納される音声発話メタデータとしては、音声発話メタデータを取得するためのアドレス情報を記述したものと、音声発話メタデータの内容そのものを記述したものの2種類が存在する。そして、音声発話メタデータに、アドレス情報を記述した場合には、音声発話メタデータの内容

は、当該アドレス情報に従い取得されるファイル（以下、「音声発話メタデータファイル」という。）に記述されていることになる。

[0060] すなわち、音声発話メタデータ生成部 112 は、アドレス情報を含む音声発話メタデータを生成し、ESG情報生成部 113 に供給した場合には、当該アドレス情報に従い取得される音声発話メタデータファイルを生成して、ストリーム生成部 114 に供給することになる。一方、音声発話メタデータがその内容を含んでいる場合、音声発話メタデータ生成部 112 は、音声発話メタデータファイルを生成する必要はないため、当該音声発話メタデータのみを、ESG情報生成部 113 に供給することになる。

[0061] ESG情報生成部 113 は、放送番組等のコンテンツを選局するためのチャンネル情報として、ESG情報を生成する。また、ESG情報生成部 113 は、音声発話メタデータ生成部 112 から供給される音声発話メタデータを、ESG情報に格納（配置）する。ESG情報生成部 113 は、音声発話メタデータを含むESG情報を、ストリーム生成部 114 に供給する。

[0062] ストリーム生成部 114 は、コンテンツ取得部 111 から供給されるコンテンツデータと、ESG情報生成部 113 から供給されるESG情報に基づいて、所定の規格に準拠したストリームを生成し、送信部 115 に供給する。

[0063] また、ESG情報生成部 113 から供給されるESG情報に含まれる音声発話メタデータがアドレス情報を含んでいる場合、ストリーム生成部 114 には、音声発話メタデータ生成部 112 から音声発話メタデータファイルが供給される。この場合、ストリーム生成部 114 は、コンテンツ取得部 111 から供給されるコンテンツデータと、音声発話メタデータ生成部 112 から供給される音声発話メタデータファイルと、ESG情報生成部 113 から供給されるESG情報に基づいて、所定の規格に準拠したストリームを生成する。

[0064] 送信部 115 は、ストリーム生成部 114 から供給されるストリームに対して、例えばデジタル変調等の処理を施して、アンテナ 116 を介して、デジタル放送信号として送信する。

[0065] なお、図 8 の送信装置 10 においては、すべての機能ブロックが、単一の

装置内に配置される必要はなく、少なくとも一部の機能ブロックが他の機能ブロックとは独立した装置として構成されるようにしてもよい。例えば、音声発話メタデータ生成部 1 1 2 や ESG 情報生成部 1 1 3 は、インターネット上のサーバの機能として提供されるようにしてもよい。その場合、送信装置 1 0 は、当該サーバから提供される音声発話メタデータや ESG 情報を取得して処理することになる。

[0066] (受信装置の構成例)

図 9 は、図 7 の受信装置の構成例を示す図である。

[0067] 図 9 において、受信装置 2 0 は、受信部 2 1 2、ストリーム分離部 2 1 3、再生部 2 1 4、表示部 2 1 5、スピーカ 2 1 6、ESG 情報取得部 2 1 7、音声発話メタデータ取得部 2 1 8、及び、TTS エンジン 2 1 9 から構成される。

[0068] 受信部 2 1 2 は、アンテナ 2 1 1 で受信されたデジタル放送信号に対して復調処理等を行い、それにより得られるストリームを、ストリーム分離部 2 1 3 に供給する。

[0069] ストリーム分離部 2 1 3 は、受信部 2 1 2 から供給されるストリームから、コンテンツデータと、ESG 情報を分離して、コンテンツデータを再生部 2 1 4 に、ESG 情報を ESG 情報取得部 2 1 7 にそれぞれ供給する。

[0070] 再生部 2 1 4 は、ストリーム分離部 2 1 3 から供給されるコンテンツデータに基づいて、コンテンツの映像を表示部 2 1 5 に表示させるとともに、コンテンツの音声をスピーカ 2 1 6 から出力させる。これにより、放送番組等のコンテンツの再生が行われる。

[0071] ESG 情報取得部 2 1 7 は、ストリーム分離部 2 1 3 から供給される ESG 情報を取得する。ESG 情報取得部 2 1 7 は、例えばユーザにより ESG 画面の表示が指示された場合、ESG 情報を再生部 2 1 4 に供給する。再生部 2 1 4 は、ESG 情報取得部 2 1 7 から供給される ESG 情報に基づいて、ESG 画面を生成し、表示部 2 1 5 に表示させる。

[0072] また、ESG 情報取得部 2 1 7 は、ESG 情報に含まれる音声発話メタデータを、音声発話メタデータ取得部 2 1 8 に供給する。音声発話メタデータ取得部

218は、ESG情報取得部217から供給される音声発話メタデータを取得する。

[0073] ここで、音声発話メタデータには、音声発話メタデータを取得するためのアドレス情報を記述したものと、音声発話メタデータの内容そのものを記述したものの2種類が存在するのは、先に述べた通りである。

[0074] すなわち、音声発話メタデータ取得部218は、音声発話メタデータにアドレス情報が含まれている場合、当該アドレス情報に基づいて、ストリーム分離部213により分離されるストリームから、音声発話メタデータファイルを取得し、そこから得られる内容を含んでいる音声発話メタデータをTTSエンジン219に供給する。一方、音声発話メタデータ取得部218は、音声発話メタデータがその内容を含んでいる場合には、当該音声発話メタデータをそのまま、TTSエンジン219に供給する。

[0075] TTSエンジン219は、音声発話メタデータ取得部218から供給される音声発話メタデータに基づいて、ユーザインターフェース等の表示情報を読み上げて、その音声を、スピーカ216から出力する。

[0076] ここでは、例えば、ESG画面が表示部215に表示されている場合において、視覚障害者に対してアクセシブルにするために、タイトルや番組情報などを読み上げるに際して、テキスト情報の読み方が一意に定まらないときなどに、TTSエンジン219は、音声発話メタデータに従い、テキスト情報が、制作者の意図した通りに読み上げられるようにする。これにより、視覚障害者が、健常者と同等の情報を得られるようになる。

[0077] なお、図9の受信装置20においては、表示部215とスピーカ216が内部に設けられている構成を示したが、表示部215とスピーカ216は、外部の別の装置として設けられるようにしてもよい。

[0078] <3. ESGの拡張による音声発話メタデータの配置>

[0079] 次に、音声発話メタデータが格納されるESG情報の詳細について説明する。
なお、ESG (Electronic Service Guide) は、携帯電話の規格策定を行う組織であるOMA (Open Mobile Alliance) によりその仕様が策定されており、音声

発話メタデータが格納されるESG情報も、OMA-BCAST (OMA - Mobile Broadcast Services Enabler Suite) に規定されたESGに準拠している。

[0080] (ESGの構造)

図10は、ESGの構造の例を示す図である。なお、図10において、フラグメント間を接続する各ラインは、接続された各フラグメントでの相互参照を意味している。

[0081] 図10において、ESGは、それぞれの目的を有するフラグメントで構成され、使用する用途に応じて、アドミニストラティブ (Administrative)、プロビジョニング (Provisioning)、コア (Core)、及び、アクセス (Access) からなる4つのグループに分けられる。

[0082] アドミニストラティブは、ESG情報を受信することができる基本情報を提供するグループである。アドミニストラティブのグループは、サービスガイド伝達ディスクリプタ (ServiceGuideDeliveryDescriptor) から構成される。サービスガイド伝達ディスクリプタは、複数のサービスガイドフラグメントを受信することができるチャンネルに関する情報、チャンネルに関するスケジューリング情報、及び、更新情報を、受信装置20に提供する。これにより、受信装置20では、必要なESG情報のみを適切な時間で受信することが可能となる。

[0083] プロビジョニングは、サービス受信に関する料金情報を提供するためのグループである。プロビジョニングのグループは、購買アイテム (Purchase Item)、購買データ (Purchase Data)、及び、購買チャンネル (Purchase Channel) から構成される。購買アイテムは、サービス又はサービスバンドルに関する料金情報を提供する。購買データは、ユーザがどのような方法を通じて料金を支払うことができるかに関する情報を提供する。購買チャンネルは、ユーザが実際にサービスを購入することができるシステムに関する情報を提供する。

[0084] なお、購買アイテム、購買データ、及び、購買チャンネルの各フラグメントには、音声発話メタデータ又はその取得先を示すアドレス情報を格納するこ

とができる。購買アイテム、購買データ、及び、購買チャネルの各フラグメントに、音声発話メタデータを格納する方法については、図 13 を参照して後述する。

[0085] コアは、サービスそれ自体に関する情報を提供するグループである。コアのグループは、サービス (Service)、スケジュール (Schedule)、及び、コンテンツ (Content) から構成される。サービスは、チャンネル・サービスの内容、及び、関連する制御情報を含むメタデータを提供する。スケジュールは、コンテンツの配信スケジュール、及び、関連する制御情報を含むメタデータを提供する。コンテンツは、サービスを構成するコンテンツの内容、及び、関連する制御情報を含むメタデータを提供する。

[0086] なお、サービス、及び、コンテンツの各フラグメントには、音声発話メタデータ又はその取得先を示すアドレス情報を格納することができる。図 11 には、サービスフラグメント (Service Fragment) の構成例が示され、図 12 には、コンテンツフラグメント (Content Fragment) の構成例が示されている。サービスとコンテンツの各フラグメントに、音声発話メタデータを格納する方法については、図 13 を参照して後述する。

[0087] アクセスは、コアのグループのサービスを受信する方法を示すサービスアクセス情報、及び、サービスを構成しているコンテンツが送信されるセッションに関する具体的な情報を提供するグループであって、受信装置 20 がサービスにアクセスすることができるようにする。アクセスのグループは、アクセス (Access)、及び、セッションディスクリプション (Session Description) から構成される。

[0088] アクセスのグループ内のアクセスは、1つのサービスに関する複数のアクセス方法を受信装置 20 に提供することにより、1つのサービスに基づいていくつかの付加的なサービスにアクセスすることができる方法を提供する。セッションディスクリプションは、1つのアクセスフラグメント (Access Fragment) で定義されたサービスアクセスが送信するサービスに関するセッション情報を提供する。

[0089] また、上述した4つのグループのほかに、プレビューデータ (Preview Data) と、インタラクティビティデータ (Interactivity Data) がある。プレビューデータは、サービスとコンテンツのためのプレビューやアイコンなどを提供する。インタラクティビティデータは、サービスやコンテンツに関するアプリケーションについてのメタデータを提供する。

[0090] なお、プレビューデータフラグメント (Preview Data Fragment) には、音声発話メタデータ又はその取得先を示すアドレス情報を格納することができる。プレビューデータフラグメントに、音声発話メタデータを格納する方法については、図14を参照して後述する。

[0091] (拡張ESGの構成例)

図13は、音声発話メタデータ又はその取得先を示すアドレス情報を格納するために拡張されたESGの構成例を示す図である。なお、図13のESGの拡張は、ESGを構成する各フラグメントのうち、サービスフラグメント (Service Fragment)、コンテンツフラグメント (Content Fragment)、購買アイテムフラグメント (Purchase Item Fragment)、購買データフラグメント (Purchase Data Fragment)、及び、購買チャネルフラグメント (Purchase Channel) が対象とされる。

[0092] すなわち、これらのフラグメントには、Name要素とDescription要素が含まれるので、Name要素とDescription要素に対して、PhoneticInfoURI要素又はPhoneticInfo要素を追加する拡張が行われるようにする。なお、これらのフラグメントのPrivateExt要素に、PhoneticInfoURI要素又はPhoneticInfo要素が追加されるようにしてもよい。

[0093] 図13において、Name要素には、コンテンツフラグメントの名称が指定される。Name要素は、PhoneticInfoURI要素、PhoneticInfo要素、及び、Type属性を、子要素として含んでいる。

[0094] PhoneticInfoURI要素には、音声発話メタデータを取得するためのアドレス情報が指定される。Type属性は、PhoneticInfoURI要素とペアで使用され、音声発話メタデータの種別を示すタイプ情報が指定される。

- [0095] このアドレス情報としては、例えば、URI (Uniform Resource Identifier) が指定される。また、例えば、音声発話メタデータファイルが、FLUTE (File Delivery over Unidirectional Transport) セッションで伝送される場合には、FLUTEセッションで伝送される音声発話メタデータファイルを取得するためのアドレス情報が指定される。なお、音声発話メタデータは、音声合成マークアップ言語である、SSML (Speech Synthesis Markup Language) により記述することができる。
- [0096] PhoneticInfo要素には、音声発話メタデータの内容そのものが記述される。例えば、この音声発話メタデータの内容は、SSMLで記述される。Type属性は、PhoneticInfo要素とペアで使用され、音声発話メタデータの種別を示すタイプ情報が指定される。
- [0097] なお、図13において、出現数 (Cardinality) であるが、“1..N”が指定された場合には、その要素又は属性は1以上指定され、“0..N”が指定された場合には、その要素又は属性を1以上指定するかどうかは任意である。また、“0..1”が指定された場合には、その要素又は属性を指定するかどうかは任意である。
- [0098] したがって、Name要素の子要素のPhoneticInfoURI要素、PhoneticInfo要素、及び、Type属性は、オプションな要素又は属性であって、PhoneticInfoURI要素とPhoneticInfo要素は、一方の要素のみが配置されるだけでなく、その両方の要素が配置されるようにしてもよい。
- [0099] また、図13において、Description要素は、PhoneticInfoURI要素、PhoneticInfo要素、及び、Type属性を、子要素として含んでいる。すなわち、Description要素の子要素は、上述したName要素の子要素と同様とされる。
- [0100] 具体的には、PhoneticInfoURI要素には、音声発話メタデータを取得するためのアドレス情報が指定される。Type属性は、PhoneticInfoURI要素とペアで使用され、音声発話メタデータの種別を示すタイプ情報が指定される。また、PhoneticInfo要素には、音声発話メタデータの内容そのものが記述される。Type属性は、PhoneticInfo要素とペアで使用され、音声発話メタデータの

種別を示すタイプ情報が指定される。

[0101] なお、Description要素の子要素であるPhoneticInfoURI要素と、PhoneticInfo要素についても、いずれか一方の要素を配置してもよいし、それらの要素の両方が配置されるようにしてもよい。

[0102] 図14は、音声発話メタデータ又はその取得先を示すアドレス情報を格納するために拡張されたESGの他の構成例を示す図である。なお、図14のESGの拡張は、ESGを構成する各フラグメントのうち、プレビューデータフラグメント（Preview Data Fragment）が対象とされる。

[0103] すなわち、プレビューデータフラグメントには、Picture要素が含まれるので、そのrelativePreference属性に対して、PhoneticInfoURI要素又はPhoneticInfo要素を追加する拡張が行われるようにする。なお、PhoneticInfoURI要素又はPhoneticInfo要素は、プレビューデータフラグメントのPrivateExt要素に追加されるようにしてもよい。

[0104] 図14において、Picture要素には、サービスとコンテンツのためのプレビューやアイコン等が定義される。Picture要素は、PhoneticInfoURI要素、PhoneticInfo要素、及び、Type属性を、子要素として含んでいる。すなわち、Picture要素の子要素は、上述したName要素やDescription要素の子要素と同様とされる。

[0105] 具体的には、PhoneticInfoURI要素には、音声発話メタデータを取得するためのアドレス情報が指定される。Type属性は、PhoneticInfoURI要素とペアで使用され、音声発話メタデータの種別を示すタイプ情報が指定される。また、PhoneticInfo要素には、音声発話メタデータの内容そのものが記述される。Type属性は、PhoneticInfo要素とペアで使用され、音声発話メタデータの種別を示すタイプ情報が指定される。

[0106] なお、Picture要素の子要素であるPhoneticInfoURI要素と、PhoneticInfo要素についても、いずれか一方の要素を配置してもよいし、それらの要素の両方が配置されるようにしてもよい。

[0107] （PhoneticInfoURI要素の構成）

図 1 5 は、拡張されたESGにおけるPhoneticInfoURI要素の詳細な構成を示す図である。

[0108] 図 1 5 において、PhoneticInfoURI要素は、サービスフラグメント等のname要素若しくはDescription要素、又は、プレビューデータフラグメントのpicture要素の子要素として記述される。PhoneticInfoURI要素には、音声発話メタデータを取得するためのアドレス情報が指定される。

[0109] また、PhoneticInfoURI要素のtype属性として、音声発話メタデータのエンコーディングフォーマットの識別URIが指定される。

[0110] 例えば、音声発話メタデータファイルは、FLUTEセッションで伝送されるので、PhoneticInfoURI要素には、FLUTEセッションで伝送される音声発話メタデータファイルを取得するためのアドレス情報が記述される。

[0111] (PhoneticInfo要素の構成)

図 1 6 は、拡張されたESGにおけるPhoneticInfo要素の詳細な構成を示す図である。

[0112] 図 1 6 において、PhoneticInfo要素は、サービスフラグメント等のname要素若しくはDescription要素、又は、プレビューデータフラグメントのpicture要素の子要素として記述される。PhoneticInfo要素には、音声発話メタデータの内容そのものが記述される。

[0113] また、PhoneticInfo要素のtype属性として、音声発話メタデータのエンコーディングフォーマットの識別URIが指定される。

[0114] 例えば、音声発話メタデータの内容は、音声合成マークアップ言語であるSMLで記述され、PhoneticInfo要素の開始タグと終了タグの間に、テキスト情報として収められる。

[0115] なお、拡張されたESGの対象のフラグメントには、PhoneticInfoURI要素及びPhoneticInfo要素の少なくとも 1 つの要素が記述される。また、音声発話メタデータは、PhoneticInfoURI要素又はPhoneticInfo要素により指定されるため、「PhoneticInfoオブジェクト」と称される場合がある。

[0116] <4. 音声発話メタデータの記述例>

[0117] 上述したように、音声発話メタデータは、例えば、音声合成マークアップ言語であるSSMLで記述することができる。このSSMLは、W3C (World Wide Web Consortium) によって、より高品質な音声合成機能を利用可能にすることを目的として勧告されたものである。SSMLを用いることで、発音や音量、調子など、音声合成に必要な要素をきめ細かく、かつ適度に制御することが可能となる。以下、図17乃至図19には、SSML形式の文書の記述例を例示している。

[0118] (sub要素)

図17は、SSML形式におけるsub要素の記述例を示す図である。

[0119] sub要素は、テキスト情報を別のテキスト情報に置き換えるために用いられる。alias属性には、音声発話用のテキスト情報が指定される。例えば、図17において、“W3C”であるテキスト情報は、“World Wide Web Consortium”である音声発話用のテキスト情報に変換されて読み上げられる。

[0120] このsub要素を用いることで、例えば、テキスト情報の読み方が一意に定まらない場合にその音声の読み方を示した情報を指定することが可能となる。

[0121] (phoneme要素)

図18は、SSML形式におけるphoneme要素の記述例を示す図である。

[0122] phoneme要素は、記述されているテキスト情報に、音素/音声上の発音を付与するために用いられる。phoneme要素は、alphabet属性と、ph属性を指定することができる。alphabet属性には、音素/音声の発音文字が指定される。ph属性には、音素/音声の文字列が指定される。例えば、図18においては、“La vita e bella”であるテキスト情報の読み方が、ph属性で指定される。なお、alphabet属性に指定された“ipa”は、国際音声記号 (IPA: International Phonetic Alphabet) の読み記号に対応していることを示している。

[0123] このphoneme要素を用いることで、例えば、発音が難解な固有名詞等である場合にその音素情報などを指定することが可能となる。

[0124] (audio要素)

図19は、SSML形式におけるaudio要素の記述例を示す図である。

[0125] audio要素は、音声ファイルの組み込み音声や合成音声を出力するために用いられる。audio要素は、src属性を指定することができる。src属性には、音声ファイルのURI (Uniform Resource Identifier) が指定される。例えば、図19においては、“What city do you want to fly from?”であるテキスト情報が、src属性に指定された“prompt.au”である音声ファイルを再生することで読み上げられる。

[0126] このaudio要素を用いることで、例えば、録音済みの音声ファイルの再生が可能となり、ユーザインターフェースの制作者が意図した通りの音声情報を視覚障害者に提供可能となる。

[0127] なお、上述したsub要素、phoneme要素、及び、audio要素は、SSML形式を用いた音声発話メタデータの記述方法の一例であって、SSML形式の他の要素や属性を用いるようにしてもよい。また、音声発話メタデータは、SSML形式以外の他のマークアップ言語などで記述するようにしてもよい。

[0128] <5. 各装置で実行される処理の流れ>

[0129] 次に、図7の放送システム1を構成する送信装置10と受信装置20で実行される処理の流れを説明する。

[0130] (送信処理)

まず、図20のフローチャートを参照して、図7の送信装置10により実行される、送信処理の流れを説明する。

[0131] ステップS111において、コンテンツ取得部111は、放送番組等のコンテンツを取得して、ストリーム生成部114に供給する。

[0132] ステップS112において、音声発話メタデータ生成部112は、例えばユーザインターフェースの制作者からの指示に従い、音声発話メタデータを生成して、ESG情報生成部113に供給する。

[0133] なお、音声発話メタデータ生成部112は、アドレス情報を含む音声発話メタデータを生成してESG情報生成部113に供給した場合には、当該アドレス情報に従い取得される音声発話メタデータファイルを生成し、ストリーム生成部114に供給する。

- [0134] ステップS 1 1 3において、ESG情報生成部 1 1 3は、音声発話メタデータ生成部 1 1 2から供給される音声発話メタデータに基づいて、ESG情報を生成し、ストリーム生成部 1 1 4に供給する。
- [0135] ステップS 1 1 4において、ストリーム生成部 1 1 4は、コンテンツ取得部 1 1 1から供給されるコンテンツデータと、ESG情報生成部 1 1 3から供給されるESG情報に基づいて、所定の規格に準拠したストリームを生成し、送信部 1 1 5に供給する。
- [0136] なお、ストリーム生成部 1 1 4は、ESG情報生成部 1 1 3から供給されるESG情報に含まれる音声発話メタデータがアドレス情報を含んでいる場合、コンテンツデータとESG情報に加えて、音声発話メタデータ生成部 1 1 2から供給される音声発話メタデータファイルに基づいて、所定の規格に準拠したストリームを生成し、送信部 1 1 5に供給する。
- [0137] ステップS 1 1 5において、送信部 1 1 5は、ストリーム生成部 1 1 4から供給されるストリームに対して、例えばデジタル変調等の処理を施して、アンテナ 1 1 6を介して、デジタル放送信号として送信する。
- [0138] 以上、送信処理について説明した。この送信処理では、ユーザインターフェース等の表示情報に対する制作者が意図する音声の発話に関する音声発話メタデータが生成され、音声発話メタデータを含むESG情報が生成され、コンテンツとともに、音声発話メタデータを含むESG情報が送信される。
- [0139] これにより、受信装置 2 0側では、TTSエンジン 2 1 9が、音声発話メタデータに基づいて、表示情報を読み上げるので、例えば、テキスト情報の読み方が一意に定まらない場合や、発音が難解な固有名詞等である場合などであっても、確実に、制作者が意図した通りにテキスト情報が読み上げられる。その結果、視覚障害者が、健常者と同等の情報を得られるようになる。
- [0140] (受信処理)
- 次に、図 2 1 のフローチャートを参照して、図 7 の受信装置 2 0により実行される、受信処理の流れを説明する。
- [0141] ステップS 2 1 1において、受信部 2 1 2は、アンテナ 2 1 1を介して、

送信装置 10 から送信されてくるデジタル放送信号を受信する。また、受信部 212 は、デジタル放送信号に対して復調処理等を行い、それにより得られるストリームを、ストリーム分離部 213 に供給する。

[0142] ステップ S 212 において、ストリーム分離部 213 は、受信部 212 から供給されるストリームから、コンテンツデータと、ESG情報を分離して、コンテンツデータを再生部 214 に、ESG情報をESG情報取得部 217 にそれぞれ供給する。

[0143] ステップ S 213 において、ESG情報取得部 217 は、ストリーム分離部 213 から供給されるESG情報を取得する。ESG情報取得部 217 は、例えばユーザによりESG画面の表示が指示された場合、ESG情報を再生部 214 に供給する。また、ESG情報取得部 217 は、ESG情報に含まれる音声発話メタデータを、音声発話メタデータ取得部 218 に供給する。

[0144] ステップ S 214 において、再生部 214 は、ESG情報取得部 217 から供給されるESG情報に基づいて、ESG画面を生成し、表示部 215 に表示させる。

[0145] ステップ S 215 において、音声発話メタデータ取得部 218 は、ESG情報取得部 217 から供給される音声発話メタデータを取得する。

[0146] ここで、音声発話メタデータ取得部 218 は、音声発話メタデータにアドレス情報が含まれている場合、当該アドレス情報に基づいて、ストリーム分離部 213 により分離されるストリームから、音声発話メタデータファイルを取得し、そこから得られる内容を含んでいる音声発話メタデータをTTSエンジン 219 に供給する。一方、音声発話メタデータ取得部 218 は、音声発話メタデータがその内容を含んでいる場合には、当該音声発話メタデータをそのまま、TTSエンジン 219 に供給する。

[0147] ステップ S 216 において、TTSエンジン 219 は、音声発話メタデータ取得部 218 から供給される音声発話メタデータに基づいて、ユーザインターフェース等の表示情報を読み上げて、その音声を、スピーカ 216 から出力する。

- [0148] ここでは、ステップS 2 1 4 の処理で、表示部 2 1 5 に、ESG画面が表示されている場合において、視覚障害者に対してアクセシブルにするために、タイトルや番組情報などを読み上げるに際して、テキスト情報の読み方が一意に定まらないときなどに、TTSエンジン 2 1 9 は、音声発話メタデータに従い、テキスト情報が、制作者の意図した通りに読み上げられるようにする。
- [0149] 以上、受信処理について説明した。この受信処理では、送信装置 1 0 から送信されてくる、表示情報に対する制作者が意図する音声の発話に関する音声発話メタデータを含むESG情報が受信され、ESG情報に含まれる音声発話メタデータが取得され、音声発話メタデータに基づいて、ユーザインターフェース等の表示情報が読み上げられる。
- [0150] これにより、TTSエンジン 2 1 9 においては、音声発話メタデータに基づいて、表示情報を読み上げるので、例えば、テキスト情報の読み方が一意に定まらない場合や、発音が難解な固有名詞等である場合などであっても、確実に、制作者が意図した通りにテキスト情報が読み上げられる。その結果、視覚障害者が、健常者と同等の情報を得られるようになる。
- [0151] < 6. 変形例 >
- [0152] 上述した説明では、音声発話メタデータにアドレス情報が含まれている場合には、当該アドレス情報に従い、FLUTEセッションで伝送される音声発話メタデータファイルが取得されるとして説明したが、音声発話メタデータファイルは、インターネット上のサーバから配信されるようにしてもよい。この場合、アドレス情報としては、サーバのURL (Uniform Resource Locator) 等が指定される。
- [0153] また、上述した説明では、電子番組情報として、OMA-BCASTで規定されたESGを説明したが、本技術は、例えば、EPG (Electronic Program Guide) その他の電子番組情報に適用することができる。さらに、ESG情報などの電子番組情報が、インターネット上のサーバから配信され、受信装置 2 0 により受信されるようにしてもよい。
- [0154] < 7. コンピュータの構成 >

- [0155] 上述した一連の処理は、ハードウェアにより実行することもできるし、ソフトウェアにより実行することもできる。一連の処理をソフトウェアにより実行する場合には、そのソフトウェアを構成するプログラムが、コンピュータにインストールされる。図22は、上述した一連の処理をプログラムにより実行するコンピュータのハードウェアの構成例を示す図である。
- [0156] コンピュータ900において、CPU (Central Processing Unit) 901, ROM (Read Only Memory) 902, RAM (Random Access Memory) 903は、バス904により相互に接続されている。バス904には、さらに、入出力インターフェース905が接続されている。入出力インターフェース905には、入力部906、出力部907、記録部908、通信部909、及び、ドライブ910が接続されている。
- [0157] 入力部906は、キーボード、マウス、マイクロフォンなどよりなる。出力部907は、ディスプレイ、スピーカなどよりなる。記録部908は、ハードディスクや不揮発性のメモリなどよりなる。通信部909は、ネットワークインターフェースなどよりなる。ドライブ910は、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリなどのリムーバブルメディア911を駆動する。
- [0158] 以上のように構成されるコンピュータ900では、CPU901が、ROM902や記録部908に記憶されているプログラムを、入出力インターフェース905及びバス904を介して、RAM903にロードして実行することにより、上述した一連の処理が行われる。
- [0159] コンピュータ900 (CPU901) が実行するプログラムは、例えば、パッケージメディア等としてのリムーバブルメディア911に記録して提供することができる。また、プログラムは、ローカルエリアネットワーク、インターネット、デジタル衛星放送といった、有線又は無線の伝送媒体を介して提供することができる。
- [0160] コンピュータ900では、プログラムは、リムーバブルメディア911をドライブ910に装着することにより、入出力インターフェース905を介

して、記録部 908 にインストールすることができる。また、プログラムは、有線又は無線の伝送媒体を介して、通信部 909 で受信し、記録部 908 にインストールすることができる。その他、プログラムは、ROM 902 や記録部 908 に、あらかじめインストールしておくことができる。

[0161] ここで、本明細書において、コンピュータがプログラムに従って行う処理は、必ずしもフローチャートとして記載された順序に沿って時系列に行われる必要はない。すなわち、コンピュータがプログラムに従って行う処理は、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含む。また、プログラムは、1 のコンピュータ（プロセッサ）により処理されるものであってもよいし、複数のコンピュータによって分散処理されるものであってもよい。

[0162] なお、本技術の実施の形態は、上述した実施の形態に限定されるものではなく、本技術の要旨を逸脱しない範囲において種々の変更が可能である。

[0163] また、本技術は、以下のような構成をとることができる。

[0164] (1)

表示情報に対する制作者が意図する音声の発話に関するメタデータを生成するメタデータ生成部と、

前記メタデータを含む電子番組情報を生成する電子番組情報生成部と、

前記表示情報を表示可能な受信装置に対して、前記電子番組情報を送信する送信部と

を備える送信装置。

(2)

前記メタデータは、読み方が一意に定まらない文字列、又は発音が難解な文字列の発話に関する情報を含んでいる

(1) に記載の送信装置。

(3)

前記表示情報は、コンテンツに関する情報、又はアイコンを含んでいる

(1) 又は (2) に記載の送信装置。

(4)

前記コンテンツを取得するコンテンツ取得部をさらに備え、

前記送信部は、前記電子番組情報を、前記コンテンツとともに、デジタル放送信号で送信する

(3)に記載の送信装置。

(5)

前記電子番組情報は、OMA-BCAST (Open Mobile Alliance - Mobile Broadcast Services Enabler Suite) で規定されたESG (Electronic Service Guide) に準拠しており、

前記メタデータは、SSML (Speech Synthesis Markup Language) 形式で記述され、

前記ESGを構成する所定のフラグメントに、前記SSML形式で記述された前記メタデータのファイルの取得先を示すアドレス情報、又は前記SSML形式で記述された前記メタデータの内容そのものが含まれる

(1)乃至(4)のいずれかに記載の送信装置。

(6)

送信装置の送信方法において、

前記送信装置が、

表示情報に対する制作者が意図する音声の発話に関するメタデータを生成し、

前記メタデータを含む電子番組情報を生成し、

前記表示情報を表示可能な受信装置に対して、前記電子番組情報を送信する

ステップを含む送信方法。

(7)

送信装置から送信されてくる、表示情報に対する制作者が意図する音声の発話に関するメタデータを含む電子番組情報を受信する受信部と、

前記電子番組情報に含まれる前記メタデータを取得するメタデータ取得部

と、

前記メタデータに基づいて、前記表示情報を読み上げる音声読み上げ部とを備える受信装置。

(8)

前記メタデータは、読み方が一意に定まらない文字列、又は発音が難解な文字列の発話に関する情報を含んでいる

(7)に記載の受信装置。

(9)

前記表示情報は、コンテンツに関する情報、又はアイコンを含んでいる

(7)又は(8)に記載の受信装置。

(10)

前記受信部は、デジタル放送信号として、前記コンテンツとともに送信される前記電子番組情報を受信する

(9)に記載の受信装置。

(11)

前記電子番組情報は、OMA-BCASTで規定されたESGに準拠しており、

前記メタデータは、SSML形式で記述され、

前記ESGを構成する所定のフラグメントに、前記SSML形式で記述された前記メタデータのファイルの取得先を示すアドレス情報、又は前記SSML形式で記述された前記メタデータの内容そのものが含まれており、

前記メタデータ取得部は、前記アドレス情報に従い前記メタデータのファイルを取得するか、又は前記フラグメントから前記メタデータを取得する

(7)乃至(10)のいずれかに記載の受信装置。

(12)

受信装置の受信方法において、

前記受信装置が、

送信装置から送信されてくる、表示情報に対する制作者が意図する音声の発話に関するメタデータを含む電子番組情報を受信し、

前記電子番組情報に含まれる前記メタデータを取得し、
前記メタデータに基づいて、前記表示情報を読み上げる
ステップを含む受信方法。

符号の説明

[0165] 1 放送システム, 10 送信装置, 20 受信装置, 111 コンテンツ取得部, 112 音声発話メタデータ生成部, 113 ESG情報生成部, 114 ストリーム生成部, 115 送信部, 212 受信部, 213 ストリーム分離部, 214 再生部, 215 表示部, 216 スピーカ, 217 ESG情報取得部, 218 音声発話メタデータ取得部, 219 TTSエンジン, 900 コンピュータ, 901 CPU

請求の範囲

- [請求項1] 表示情報に対する制作者が意図する音声の発話に関するメタデータを生成するメタデータ生成部と、
 前記メタデータを含む電子番組情報を生成する電子番組情報生成部と、
 前記表示情報を表示可能な受信装置に対して、前記電子番組情報を送信する送信部と
 を備える送信装置。
- [請求項2] 前記メタデータは、読み方が一意に定まらない文字列、又は発音が難解な文字列の発話に関する情報を含んでいる
 請求項1に記載の送信装置。
- [請求項3] 前記表示情報は、コンテンツに関する情報、又はアイコンを含んでいる
 請求項2に記載の送信装置。
- [請求項4] 前記コンテンツを取得するコンテンツ取得部をさらに備え、
 前記送信部は、前記電子番組情報を、前記コンテンツとともに、デジタル放送信号で送信する
 請求項3に記載の送信装置。
- [請求項5] 前記電子番組情報は、OMA-BCAST (Open Mobile Alliance - Mobile Broadcast Services Enabler Suite) で規定されたESG (Electronic Service Guide) に準拠しており、
 前記メタデータは、SSML (Speech Synthesis Markup Language) 形式で記述され、
 前記ESGを構成する所定のフラグメントに、前記SSML形式で記述された前記メタデータのファイルの取得先を示すアドレス情報、又は前記SSML形式で記述された前記メタデータの内容そのものが含まれる
 請求項1に記載の送信装置。
- [請求項6] 送信装置の送信方法において、

前記送信装置が、
表示情報に対する制作者が意図する音声の発話に関するメタデータを生成し、
前記メタデータを含む電子番組情報を生成し、
前記表示情報を表示可能な受信装置に対して、前記電子番組情報を送信する
ステップを含む送信方法。

[請求項7] 送信装置から送信されてくる、表示情報に対する制作者が意図する音声の発話に関するメタデータを含む電子番組情報を受信する受信部と、
前記電子番組情報に含まれる前記メタデータを取得するメタデータ取得部と、
前記メタデータに基づいて、前記表示情報を読み上げる音声読み上げ部と
を備える受信装置。

[請求項8] 前記メタデータは、読み方が一意に定まらない文字列、又は発音が難解な文字列の発話に関する情報を含んでいる
請求項7に記載の受信装置。

[請求項9] 前記表示情報は、コンテンツに関する情報、又はアイコンを含んでいる
請求項8に記載の受信装置。

[請求項10] 前記受信部は、デジタル放送信号として、前記コンテンツとともに送信される前記電子番組情報を受信する
請求項9に記載の受信装置。

[請求項11] 前記電子番組情報は、OMA-BCASTで規定されたESGに準拠しており、
前記メタデータは、SSML形式で記述され、
前記ESGを構成する所定のフラグメントに、前記SSML形式で記述された前記メタデータのファイルの取得先を示すアドレス情報、又は前

記SSML形式で記述された前記メタデータの内容そのものが含まれており、

前記メタデータ取得部は、前記アドレス情報に従い前記メタデータのファイルを取得するか、又は前記フラグメントから前記メタデータを取得する

請求項7に記載の受信装置。

[請求項12]

受信装置の受信方法において、

前記受信装置が、

送信装置から送信されてくる、表示情報に対する制作者が意図する音声の発話に関するメタデータを含む電子番組情報を受信し、

前記電子番組情報に含まれる前記メタデータを取得し、

前記メタデータに基づいて、前記表示情報を読み上げる

ステップを含む受信方法。