

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7601224号
(P7601224)

(45)発行日 令和6年12月17日(2024.12.17)

(24)登録日 令和6年12月9日(2024.12.9)

(51)国際特許分類

F I

G 1 0 L 13/06 (2013.01)

G 1 0 L 13/06 1 4 0

G 1 0 L 25/30 (2013.01)

G 1 0 L 25/30

請求項の数 7 (全27頁)

(21)出願番号	特願2023-532858(P2023-532858)	(73)特許権者	000004226
(86)(22)出願日	令和3年7月5日(2021.7.5)		日本電信電話株式会社
(86)国際出願番号	PCT/JP2021/025237		東京都千代田区大手町一丁目5番1号
(87)国際公開番号	WO2023/281555	(74)代理人	110002147
(87)国際公開日	令和5年1月12日(2023.1.12)		弁理士法人酒井国際特許事務所
審査請求日	令和5年11月1日(2023.11.1)	(72)発明者	金川 裕紀
			東京都千代田区大手町一丁目5番1号
			日本電信電話株式会社内
		審査官	山下 剛史

最終頁に続く

(54)【発明の名称】 生成方法、生成プログラム及び生成装置

(57)【特許請求の範囲】

【請求項1】

音声波形の音響特徴量を基にして、サブバンド信号の中間表現情報を生成する中間表現生成工程と、

前記サブバンド信号のチャンネル数と、同時生成するサンプル数との数に応じた複数の確率分布生成モデルであって、それぞれに割り当てられた時刻および帯域に対応するサブバンド信号の情報を出力する前記複数の確率分布生成モデルに、前記中間表現情報を入力することで、異なる複数の時刻および異なる複数の帯域に応じた複数のサブバンド信号を同時に生成するサブバンド信号生成工程と、

前記複数のサブバンド信号を基にして、音声波形を生成する音声波形生成工程とを含んだことを特徴とする生成方法。

10

【請求項2】

前記音響特徴量を入力した場合に、前記音響特徴量の中間表現情報を出力する第1中間表現モデルを用いて、前記音声波形の音響特徴量を、前記音響特徴量の中間表現情報に変換する変換工程を更に含んだことを特徴とする請求項1に記載の生成方法。

【請求項3】

前記中間表現生成工程は、前記音響特徴量の中間表現情報を入力した場合に、前記サブバンド信号の中間表現情報を出力する第2中間表現モデルを用いて、前記サブバンド信号の中間表現情報を生成することを特徴とする請求項2に記載の生成方法。

【請求項4】

20

音声波形から算出される複数のサブバンド信号と、前記サブバンド信号生成工程によって生成される複数のサブバンド信号とを基にして、損失値を計算し、前記損失値を基にして、前記第 1 中間表現モデル、前記第 2 中間表現モデル、前記複数の確率分布生成モデルのうち少なくとも 1 つのモデルの学習を実行する学習工程を更に有することを特徴とする請求項 3 に記載の生成方法。

【請求項 5】

前記サブバンド信号生成工程は、1 つのモデルから複数の時間帯、および、複数の帯域に対応するサブバンド信号の情報を同時に出力する同時確率分布生成モデルを用いて、前記複数のサブバンド信号を同時に生成することを特徴とする請求項 1 に記載の生成方法。

【請求項 6】

音声波形の音響特徴量を基にして、サブバンド信号の中間表現情報を生成する中間表現生成ステップと、

前記サブバンド信号のチャンネル数と、同時生成するサンプル数との数に応じた複数の確率分布生成モデルであって、それぞれに割り当てられた時刻および帯域に対応するサブバンド信号の情報を出力する前記複数の確率分布生成モデルに、前記中間表現情報を入力することで、異なる複数の時刻および異なる複数の帯域に応じた複数のサブバンド信号を同時に生成するサブバンド信号生成ステップと、

前記複数のサブバンド信号を基にして、音声波形を生成する音声波形生成ステップとをコンピュータに実行させるための生成プログラム。

【請求項 7】

音声波形の音響特徴量を基にして、サブバンド信号の中間表現情報を生成する中間表現生成し、前記サブバンド信号のチャンネル数と、同時生成するサンプル数との数に応じた複数の確率分布生成モデルであって、それぞれに割り当てられた時刻および帯域に対応するサブバンド信号の情報を出力する前記複数の確率分布生成モデルに、前記中間表現情報を入力することで、異なる複数の時刻および異なる複数の帯域に応じた複数のサブバンド信号を同時に生成する生成部と、

前記複数のサブバンド信号を基にして、音声波形を生成する音声波形生成部とを備えることを特徴とする生成装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、生成方法、生成プログラム及び生成装置に関する。

【背景技術】

【0002】

音声合成において、スペクトルや声の高さを表すピッチ等の音響特徴量から音声波形に変換するモジュールはボコーダーと呼ばれる。ボコーダーの実装方法は大きく二種類がある。一つは信号処理による方法であり、STRAIGHTやWORLDといった手法が有名である（たとえば、非特許文献 1，2 参照）。これらの方法は数理モデルにより音響特徴量から音声波形への変換を表現するため、学習が不要かつ処理速度が高速であるが、分析再合成された音声を自然音声と比較すると品質が劣る。

【0003】

もう一つは、WaveNetに代表されるニューラルネットによる手法（ニューラルボコーダー）が提案されている（たとえば、特許文献 1 参照）。こちらは自然音声と比較しても遜色ない品質の音声を合成可能な一方、計算量が多いため信号処理のボコーダーよりも動作が低速である。通常、1 つの音声サンプルの予測のためにニューラルネットを 1 回順伝搬しなくてはならないため、そのまま実装したのではリアルタイム動作は困難である。

【0004】

ニューラルボコーダーの計算量を減らし、特に CPU においてリアルタイム動作させるためには主に二つのアプローチが採られる。一つはニューラルネットの順伝搬 1 回あたりの計算コストを削減するもので、WaveNet で用いられる巨大な畳み込みニューラルネット(C

10

20

30

40

50

NN: Convolutional Neural Network)を小規模な再帰型ニューラルネット(RNN: Recurrent Neural Network)で置き換えたWaveRNN(たとえば、特許文献2)や、音声波形の生成過程に信号処理の知見である線形予測分析(LPC)を活用したLPCNet等(たとえば、非特許文献3)がある。

【0005】

もう一つは順伝搬の回数そのものを減らす方法であり、予測対象を音声波形でなく、サブバンド信号と呼ばれる疑似直交ミラーフィルタ(Pseudo-QMF)を用いて帯域分割した短い系列に変更することで順伝搬の回数を減らせることが開示されている(たとえば、非特許文献4)。

【先行技術文献】

10

【特許文献】

【0006】

【文献】国際公開第2018/048934号

【文献】国際公開第2019/155054号

【非特許文献】

【0007】

【文献】Hideki Kawahara, Ikuyo Masuda-Katsuse and Alain de Cheveigne, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Communication, vol. 27, no. 3-4, pp. 187-207, 1999.

20

【文献】Masanori Morise, Fumiya Yokomori, Kenji Ozawa, "WORLD: a Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications," IEICE transactions on information and systems, vol. E99-D, no. 7, pp. 1877-1884, 2016.

【文献】Jean-Marc Valin and Jan Skoglund, "LPCNET: IMPROVING NEURAL SPEECH SYNTHESIS THROUGH LINEAR PREDICTION," Proc. ICASSP, 2019, pp. 5891-5895

【文献】Chengzhu Yu, Heng Lu, Na Hu, Meng Yu, Chao Weng, Kun Xu, Peng Liu, Deyi Tuo, Shiyin Kang, Guangzhi Lei, Dan Su, Dong Yu, "DURIAN: Duration Informed Attention Network for Speech Synthesis", Proc. INTERSPEECH 2020, pp. 2027-2031, 2020

30

【発明の概要】

【発明が解決しようとする課題】

【0008】

非特許文献4では引用文献2を改良し、1サンプル分の音声波形の代わりに1サンプル分のサブバンド信号を使い、次の時刻のサブバンド信号を予測する。この場合にB個のチャンネルでサブバンド表現すると系列長を1/Bに短縮でき、純伝搬回数を1/Bに削減可能である。しかしRNN等の自己回帰構造に基づき、過去のサンプルに基づき1サンプルずつ生成しているため、最大B倍速以上の高速化ができない。

40

【0009】

本発明は、上記に鑑みてなされたものであって、音響特徴量から高速に音声波形を生成することができる生成方法、生成プログラム及び生成装置を提供することを目的とする。

【課題を解決するための手段】

【0010】

上述した課題を解決し、目的を達成するために、コンピュータは、音声波形の音響特徴量を基にして、サブバンド信号の中間表現情報を生成する中間表現生成工程と、サブバンド信号のチャンネル数と、同時生成するサンプル数との数に応じた複数の確率分布生成モデルであって、それぞれに割り当てられた時刻および帯域に対応するサブバンド信号の情報を出力する複数の確率分布生成モデルに、中間表現情報を入力することで、異なる複数

50

の時刻および異なる複数の帯域に応じた複数のサブバンド信号を同時に生成するサブバンド信号生成工程と、複数のサブバンド信号を基にして、音声波形を生成する音声波形生成工程とを含む。

【発明の効果】

【0011】

本発明によれば、音響特徴量から高速に音声波形を生成することができる。

【図面の簡単な説明】

【0012】

【図1】図1は、本実施例1に係る生成装置の構成を示す機能ブロック図である。

【図2】図2は、サブバンド信号の一例を示す図である。

10

【図3】図3は、本実施例1に係る学習部を説明するための図である。

【図4】図4は、本実施例1に係る生成部を説明するための図である。

【図5】図5は、本実施例1に係る生成装置の学習時の処理手順を示すフローチャートである。

【図6】図6は、本実施例1に係る生成装置の生成時の処理手順を示すフローチャートである。

【図7】図7は、本実施例2に係る生成装置の構成を示す機能ブロック図である。

【図8】図8は、本実施例2に係る学習部を説明するための図である。

【図9】図9は、本実施例2に係る生成部を説明するための図である。

【図10】図10は、本実施例2に係る生成装置の学習時の処理手順を示すフローチャートである。

20

【図11】図11は、本実施例2に係る生成装置の生成時の処理手順を示すフローチャートである。

【図12】図12は、本実施例3に係る生成装置の構成を示す機能ブロック図である。

【図13】図13は、本実施例3に係る学習部を説明するための図である。

【図14】図14は、本実施例3に係る生成部を説明するための図である。

【図15】図15は、本実施例3に係る生成装置の学習時の処理手順を示すフローチャートである。

【図16】図16は、本実施例3に係る生成装置の生成時の処理手順を示すフローチャートである。

30

【図17】図17は、生成プログラムを実行するコンピュータの一例を示す図である。

【発明を実施するための形態】

【0013】

以下に、本願の開示する生成方法、生成プログラム及び生成装置の実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの発明が限定されるものではない。

【実施例1】

【0014】

まず、本実施例1に係る生成装置の構成例について説明する。図1は、本実施例1に係る生成装置の構成を示す機能ブロック図である。図1に示すように、この生成装置100は、通信制御部110と、入力部120と、出力部130と、記憶部140と、制御部150とを有する。

40

【0015】

通信制御部110は、NIC (Network Interface Card) 等で実現され、LAN (Local Area Network) やインターネットなどの電気通信回線を介した外部の装置と制御部150との通信を制御する。

【0016】

入力部120は、キーボードやマウス等の入力デバイスを用いて実現され、操作者による入力操作に対応して、制御部150に対して処理開始などの各種指示情報を入力する。

【0017】

出力部130は、制御部150から取得した情報を出力する出力デバイスであり、液晶

50

ディスプレイなどの表示装置、プリンター等の印刷装置等によって実現される。

【0018】

記憶部140は、学習用音声波形データ141と、第1中間表現モデル10と、第2中間表現モデル11と、確率分布生成モデル12とを有する。記憶部140は、RAM(Random Access Memory)、フラッシュメモリ(Flash Memory)等の半導体メモリ素子、または、ハードディスク、光ディスク等の記憶装置によって実現される。

【0019】

学習用音声波形データ141には、学習用の複数の音声波形のデータが含まれる。学習用音声波形データ141は、第1中間表現モデル10、第2中間表現モデル11、確率分布生成モデル12を学習する場合に利用される。

10

【0020】

第1中間表現モデル10は、音声波形から計算される音響特徴量が入力された場合に、音響特徴量の中間表現 m を出力するモデルである。第1中間表現モデル10は、たとえば、NN(Neural Network)、CNN、RNN、WaveRNN等に対応する。以下の説明では、音響特徴量の中間表現 m を、適宜、「中間表現 m 」と表記する。

【0021】

音響特徴量は、音声波形のスペクトログラム、メルケプストラム等のスペクトル情報、基本周波数、ピッチ周波数等の韻律情報に対応する。

【0022】

音響特徴量の中間表現は、音響特徴量の系列長を音声サンプル数と同じになるように伸長した情報である。係る中間表現 m を生成する第1中間表現モデル10の実現方法として、1フレームに対応する音響特徴量のベクトルをサンプル数だけ並べて伸長する方法がある。また、第1中間表現モデル10をWaveRNNで実現する場合、前後フレームの連続性を考慮するため、一次元CNNや二次元CNNを用いて特徴量変換しながら伸長する方法を採用する。

20

【0023】

第2中間表現モデル11は、中間表現 m が入力された場合に、サブバンド信号の中間表現 x_t を出力するモデルである。後述するように、第2中間表現モデル11は、複数存在する。以下の説明では、サブバンド信号の中間表現 x_t を、適宜、「中間表現 x_t 」と表記する。中間表現 x_t の t は時刻インデックスである。たとえば、 $x_{t=-1,0}$ は、時刻 $t=-1$ 、 $t=0$ の情報を含んだベクトルであることを意味する。ここで、中間表現 $x_{t=-1,0}$ は、中間表現 x_t の初期値に相当する情報である。

30

【0024】

図2は、サブバンド信号の一例を示す図である。図2の縦軸はAmplitude Responseに対応し、横軸はNormalized Frequencyに対応する。図2では、音声信号(フルバンド信号)にフィルターをかけることで、4つのサブバンド信号 s_{ub1} 、 s_{ub2} 、 s_{ub3} 、 s_{ub4} を生成した場合を示している。サブバンド信号 s_{ub1} は、低域のサブバンド信号である。サブバンド信号 s_{ub2} は、低域～中域のサブバンド信号である。サブバンド信号 s_{ub3} は、中域～高域のサブバンド信号である。サブバンド信号 s_{ub4} は、高域のサブバンド信号である。

40

【0025】

確率分布生成モデル12は、中間表現 x_t が入力された場合に、サブバンド信号の確率分布 $y_{t,b}$ を出力するモデルである。後述するように、確率分布生成モデル12は、複数存在する。確率分布 $y_{t,b}$ の t は時刻インデックスであり、 b はサブバンド信号のチャンネルのインデックスである。確率分布 $y_{t,b}$ は、音声波形サンプル値の生成確率である。音声波形が μ -law信号等の離散値で表現されている場合には、確率分布 $y_{t,b}$ として、カテゴリカル分布を用いることができる。音声波形が連続値で表現されている場合には、確率分布 $y_{t,b}$ として、ガウス分布、ベータ分布、混合ロジスティック分布を用いることができる。

【0026】

50

図 1 において、制御部 150 は、学習部 151 と、生成部 152 とを有する。制御部 150 は、CPU (Central Processing Unit) 等に対応する。

【0027】

学習部 151 は、学習用音声波形データ 141 を用いて、第 1 中間表現モデル 10、第 2 中間表現モデル 11、確率分布生成モデル 12 の学習を実行する。

【0028】

生成部 152 は、学習済みの第 1 中間表現モデル 10、第 2 中間表現モデル 11、確率分布生成モデル 12 を用いて、音響特徴量から音声波形を生成する。生成部 152 は、音響特徴量の情報を、入力部 120 から取得してもよいし、通信制御部 110 を介して、外部装置から取得してもよい。

10

【0029】

次に、図 1 に示した学習部 151 について具体的に説明する。図 3 は、本実施例 1 に係る学習部を説明するための図である。図 3 に示すように、学習部 151 は、音響特徴量計算部 20、第 1 中間表現モデル 10、複数の第 2 中間表現モデル 11 - n ($n = 1 \sim T/B - 1$)、複数の確率分布生成モデル 12 - t - b ($t = 1, \dots, T/B, b = 1, \dots, B$)、サンプリング部 21 - t - b、サブバンド信号計算部 30、損失計算部 31、モデル学習部 32 を有する。ここでは、音声の系列長を T 、サブバンド信号のチャンネル数を B 、サブバンド信号のサンプル数を $N (= 2)$ とする。

【0030】

学習部 151 は、学習用音声波形データ 141 に格納された音声波形 D_1 を音響特徴量計算部 20 に入力する。音響特徴量計算部 20 は、音声波形 D_1 を基にして、音響特徴量 D_2 を算出する。音響特徴量計算部 20 は、音響特徴量 D_2 を、第 1 中間表現モデル 10 に入力する。

20

【0031】

第 1 中間表現モデル 10 は、音響特徴量 D_2 が入力された場合に、音響特徴量の中間表現 m を出力する。第 1 中間表現モデル 10 は、中間表現 m を、第 2 中間表現モデル 11 - 1 $\sim$ 11 - n に出力する。第 1 中間表現モデル 10 は、勾配情報を、モデル学習部 32 に出力する。

【0032】

各第 2 中間表現モデル 11 - 1 $\sim$ 11 - n は、中間表現 m が入力された場合に、サブバンド信号の中間表現 x_t を出力する。たとえば、第 2 中間表現モデル 11 - 1 は、中間表現 $x_{t=1,0}$ を算出し、確率分布生成モデル 12 - 1 - 1 $\sim$ 12 - 2 - B に出力する。なお、第 2 中間表現モデル 11 - 2 $\sim$ 11 - n は、サブバンド信号の中間表現 x_t を計算する場合に、前の時刻のサブバンド信号 $s_{t,b}$ を更に用いる。サブバンド信号 $s_{t,b}$ の t は時刻インデックスであり、 b はサブバンド信号のチャンネルのインデックスである。

30

【0033】

第 2 中間表現モデル 11 - 2 は、中間表現 $x_{t=1,2}$ を算出し、確率分布生成モデル 12 - 3 - 1 $\sim$ 12 - 4 - B に出力する。第 2 中間表現モデル 11 - 3 は、中間表現 $x_{t=3,4}$ を算出し、確率分布生成モデル 12 - 5 - 1 $\sim$ 12 - 6 - B (図示略) に出力する。

【0034】

以下の説明では、第 2 中間表現モデル 11 - 1 $\sim$ 11 - n を区別しない場合、まとめて、第 2 中間表現モデル 11 と表記する。第 2 中間表現モデル 11 は、勾配情報を、モデル学習部 32 に出力する。

40

【0035】

各確率分布生成モデル 12 - t - b は、中間表現 x_t が入力された場合に、確率分布 $y_{t,b}$ を出力する。たとえば、確率分布生成モデル 12 - 1 - 1 は、確率分布 $y_{t=1,b=1}$ を、サンプリング部 21 - 1 - 1 に出力する。図示を省略するが、確率分布生成モデル 12 - 1 - 2 $\sim$ 12 - 1 - (B - 1) は、確率分布 $y_{t,b}$ を配下のサンプリング部 21 - 1 - 2 $\sim$ 21 - 1 - (B - 1) に出力する。確率分布生成モデル 12 - 1 - B は、確率分布 $y_{t=1,b=B}$ を、サンプリング部 21 - 1 - B に出力する。

50

【0036】

確率分布生成モデル12-2-1は、確率分布 $y_t = 2$ 、 $b = 1$ を、サンプリング部21-2-1に出力する。図示を省略するが、確率分布生成モデル12-2-2～12-2-(B-1)は、確率分布 y_t 、 b を配下のサンプリング部21-2-2～21-2-(B-1)に出力する。確率分布生成モデル12-2-Bは、確率分布 $y_t = 2$ 、 $b = B$ を、サンプリング部21-2-Bに出力する。

【0037】

確率分布生成モデル12-3-1は、確率分布 $y_t = 3$ 、 $b = 1$ を、サンプリング部21-3-1に出力する。図示を省略するが、確率分布生成モデル12-3-2～12-3-(B-1)は、確率分布 y_t 、 b を配下のサンプリング部21-3-2～21-3-(B-1)に出力する。確率分布生成モデル12-3-Bは、確率分布 $y_t = 3$ 、 $b = B$ を、サンプリング部21-3-Bに出力する。

10

【0038】

確率分布生成モデル12-4-1は、確率分布 $y_t = 4$ 、 $b = 1$ を、サンプリング部21-4-1に出力する。図示を省略するが、確率分布生成モデル12-4-2～12-4-(B-1)は、確率分布 y_t 、 b を配下のサンプリング部21-4-2～21-4-(B-1)に出力する。確率分布生成モデル12-4-Bは、確率分布 $y_t = 3$ 、 $b = B$ を、サンプリング部21-4-Bに出力する。

【0039】

その他の確率分布生成モデルの説明を省略するが、それぞれ、中間表現 x_t が入力された場合に、確率分布 y_t 、 b を、配下のサンプリング部に出力する。以下の説明では、各確率分布生成モデル12-t-bを区別しない場合、まとめて、確率分布生成モデル12と表記する。確率分布生成モデル12は、勾配情報を、モデル学習部32に出力する。確率分布生成モデル12は、確率分布 $y_t = 1 \sim T/B$ 、 $b = 1 \sim B$ を、損失計算部31に出力する。

20

【0040】

各サンプリング部21-t-bは、確率分布 y_t 、 b が入力された場合に、サブバンド信号 s_t 、 b を算出する。なお、学習部151は、各確率分布生成モデル12-t-bが、直接サブバンド信号を予測する場合には、各サンプリング部21-t-bを用いなくともよい。

30

【0041】

サンプリング部21-1-1は、サブバンド信号 $s_t = 1$ 、 $b = 1$ を算出する。図示を省略するが、サンプリング部21-1-2～21-1-(B-1)は、該当するサブバンド信号 s_t 、 b を算出する。サンプリング部21-1-Bは、サブバンド信号 $s_t = 1$ 、 $b = B$ を算出する。

【0042】

サンプリング部21-2-1は、サブバンド信号 $s_t = 2$ 、 $b = 1$ を算出する。図示を省略するが、サンプリング部21-2-2～21-2-(B-1)は、該当するサブバンド信号 s_t 、 b を算出する。サンプリング部21-2-Bは、サブバンド信号 $s_t = 2$ 、 $b = B$ を算出する。

40

【0043】

上記のように、サンプリング部21-1-1～21-1-Bによって、サブバンド信号 $s_t = 1$ 、 $b = 1 \sim B$ が算出される。サンプリング部21-2-1～21-2-Bによって、サブバンド信号 $s_t = 2$ 、 $b = 1 \sim B$ が算出される。サブバンド信号 $s_t = 1$ 、 $b = 1 \sim B$ 、サブバンド信号 $s_t = 2$ 、 $b = 1 \sim B$ は、第2中間表現モデル11-2に入力される。

【0044】

すなわち、確率分布生成モデル12-1-1～12-1-Bとサンプリング部21-1-1～21-1-Bとの組、及び、確率分布生成モデル12-2-1～12-2-Bとサンプリング部21-2-1～21-2-Bとの組によって、1つの中間表現 $x_t = -1, 0$ から、 $B \times N$ ($N = 2$) 個に相当するサブバンド信号 ($s_t = 1$ 、 $b = 1 \sim B$ 、 $s_t = 2$ 、 b

50

$= 1 \sim B$) が生成される。

【 0 0 4 5 】

サンプリング部 2 1 - 3 - 1 は、サブバンド信号 $s_t = 3$ 、 $b = 1$ を算出する。図示を省略するが、サンプリング部 2 1 - 3 - 2 ~ 2 1 - 3 - ($B - 1$) は、該当するサブバンド信号 s_t 、 b を算出する。サンプリング部 2 1 - 3 - B は、サブバンド信号 $s_t = 3$ 、 $b = B$ を算出する。

【 0 0 4 6 】

サンプリング部 2 1 - 4 - 1 は、サブバンド信号 $s_t = 4$ 、 $b = 1$ を算出する。図示を省略するが、サンプリング部 2 1 - 4 - 2 ~ 2 1 - 4 - ($B - 1$) は、該当するサブバンド信号 s_t 、 b を算出する。サンプリング部 2 1 - 4 - B は、サブバンド信号 $s_t = 4$ 、 $b = B$ を算出する。

10

【 0 0 4 7 】

上記のように、サンプリング部 2 1 - 3 - 1 ~ 2 1 - 3 - B によって、サブバンド信号 $s_t = 3$ 、 $b = 1 \sim B$ が算出される。サンプリング部 2 1 - 4 - 1 ~ 2 1 - 4 - B によって、サブバンド信号 $s_t = 4$ 、 $b = 1 \sim B$ が算出される。サブバンド信号 $s_t = 3$ 、 $b = 1 \sim B$ 、サブバンド信号 $s_t = 4$ 、 $b = 1 \sim B$ は、第 2 中間表現モデル 1 1 - 3 に入力される。

【 0 0 4 8 】

すなわち、確率分布生成モデル 1 2 - 3 - 1 ~ 1 2 - 3 - B とサンプリング部 2 1 - 3 - 1 ~ 2 1 - 3 - B との組、及び、確率分布生成モデル 1 2 - 4 - 1 ~ 1 2 - 4 - B とサンプリング部 2 1 - 4 - 1 ~ 2 1 - 4 - B との組によって、1 つの中間表現 $x_t = 1, 2$ から、 $B \times N$ ($N = 2$) 個に相当するサブバンド信号 ($s_t = 3$ 、 $b = 1 \sim B$ 、 $s_t = 4$ 、 $b = 1 \sim B$) が生成される。

20

【 0 0 4 9 】

その他のサンプリング部の説明を省略するが、それぞれ、確率分布 y_t 、 b が入力された場合に、サブバンド信号 s_t 、 b を算出する。各サンプリング部 2 1 - t - b から、サブバンド信号 $s_t = 1 \sim T / B$ 、 $b = 1 \sim B$ が出力される。

【 0 0 5 0 】

続いて、サブバンド信号計算部 3 0 の説明に移行する。サブバンド信号計算部 3 0 は、音声波形 D_1 から、サブバンド信号 $s_t = 1 \sim T / B$ 、 $b = 1 \sim B$ を計算する。サブバンド信号計算部 3 0 として、直交ミラーフィルタ (QMF : Quadrature Mirror Filter) や、疑似直交ミラーフィルタ (PQMF : Pseudo-Quadrature Mirror Filters) を用いることができる。直交ミラーフィルタまたは疑似直交ミラーフィルタを用いることで、入力された音声波形 D_1 の系列長を $1 / B$ に短縮したサブバンド信号に変換する。

30

【 0 0 5 1 】

サブバンド信号計算部 3 0 は、サブバンド信号 $s_t = 1 \sim T / B$ 、 $b = 1 \sim B$ を、損失計算部 3 1 に出力する。

【 0 0 5 2 】

損失計算部 3 1 は、サブバンド信号 $s_t = 1 \sim T / B$ 、 $b = 1 \sim B$ と、確率分布 $y_t = 1 \sim T / B$ 、 $b = 1 \sim B$ とを基にして、損失値 D_3 を算出する。たとえば、損失値 D_3 は、確率分布 $y_t = 1 \sim T / B$ 、 $b = 1 \sim B$ から予測されるサブバンド信号と、サブバンド信号 $s_t = 1 \sim T / B$ 、 $b = 1 \sim B$ との誤差を示す値となる。

40

【 0 0 5 3 】

損失計算部 3 1 は、確率分布にカテゴリカル分布を使う場合はクロスエントロピーを利用して損失値 D_3 を算出する。損失計算部 3 1 は、確率分布にガウス分布、ベータ分布、混合ロジスティック分布を使う場合には、負の対数尤度を利用して、損失値 D_3 を算出する。

【 0 0 5 4 】

損失計算部 3 1 は、損失値 D_3 を、モデル学習部 3 2 に出力する。

【 0 0 5 5 】

モデル学習部 3 2 は、損失値 D_3 、第 1 中間表現モデル 1 0 の勾配情報、第 2 中間表現

50

モデル 11 の勾配情報、確率分布生成モデル 12 の勾配情報を取得し、損失値 D_3 が小さくなるように、第 1 中間表現モデル 10 のパラメータ、第 2 中間表現モデル 11 のパラメータ、確率分布生成モデル 12 のパラメータを機械学習する。たとえば、モデル学習部 32 は、誤差逆伝播アルゴリズムを利用して、機械学習を実行する。

【0056】

学習部 151 に含まれる各処理部は、学習用音声波形データ 141 に含まれる複数の音声波形 D_1 について、上記処理を繰り返し実行することで、学習済みの第 1 中間表現モデル 10、第 2 中間表現モデル 11、確率分布生成モデル 12 を得る。

【0057】

次に、図 1 に示した生成部 152 について具体的に説明する。図 4 は、本実施例 1 に係る生成部を説明するための図である。図 4 に示すように、生成部 152 は、第 1 中間表現モデル 10、複数の第 2 中間表現モデル 11 - n ($n = 1 \sim T/B - 1$)、複数の確率分布生成モデル 12 - $t - b$ ($t = 1, \dots, T/B, b = 1, \dots, B$)、サンプリング部 21 - $t - b$ 、音声波形変換部 40 を有する。ここでは、音声の系列長を T 、サブバンド信号のチャンネル数を B 、サブバンド信号のサンプル数を $N (= 2)$ とする。

10

【0058】

図 3 の場合と同様にして、第 2 中間表現モデル 11 - $1 \sim 11 - n$ を区別しない場合、まとめて、第 2 中間表現モデル 11 と表記する。確率分布生成モデル 12 - $t - b$ を区別しない場合、まとめて、確率分布生成モデル 12 と表記する。

【0059】

20

第 1 中間表現モデル 10、第 2 中間表現モデル 11、確率分布生成モデル 12 は、学習部 151 によって学習済みとする。

【0060】

生成部 152 は、音響特徴量 D_{10} を、第 1 中間表現モデル 10 に入力する。第 1 中間表現モデル 10 は、音響特徴量 D_{10} が入力された場合に、音響特徴量の中間表現 m を出力する。第 1 中間表現モデル 10 は、中間表現 m を、第 2 中間表現モデル 11 - $1 \sim 11 - n$ に出力する。

【0061】

第 2 中間表現モデル 11 は、中間表現 m が入力された場合に、サブバンド信号の中間表現 x_t を、確率分布生成モデル 12 に出力する。第 2 中間表現モデル 11 に関する説明は、図 3 で行った、各第 2 中間表現モデル 11 - $1 \sim 11 - n$ に関する説明と同様である。

30

【0062】

確率分布生成モデル 12 は、中間表現 x_t が入力された場合に、確率分布 $y_{t, b}$ を、各サンプリング部 21 - $t - b$ に出力する。確率分布生成モデル 12 に関する説明は、図 3 で行った、各確率分布生成モデル 12 - $t - b$ に関する説明と同様である。

【0063】

サンプリング部 21 - $t - b$ は、確率分布 $y_{t, b}$ が入力された場合に、サブバンド信号 $s_{t, b}$ を算出する。各サンプリング部 21 - $t - b$ は、サブバンド信号 $s_{t = 1 \sim T/B, b = 1 \sim B}$ を、音声波形変換部 40 に出力する。

【0064】

40

音声波形変換部 40 は、サブバンド信号 $s_{t = 1 \sim T/B, b = 1 \sim B}$ を基にして、音声波形 $z_{t = 1 \sim T}$ を算出する。音声波形変換部 40 は、直交ミラーフィルタや、疑似直交ミラーフィルタの逆変換によって、サブバンド信号を音声波形に変換する。

【0065】

次に、本実施例 1 に係る生成装置 100 の処理手順の一例について説明する。図 5 は、本実施例 1 に係る生成装置の学習時の処理手順を示すフローチャートである。図 5 に示すように、生成装置 100 の学習部 151 は、学習用音声波形データ 141 から、音声波形を取得する (ステップ S101)。

【0066】

学習部 151 の音響特徴量計算部 20 は、音声波形から音響特徴量を計算する (ステッ

50

プ S 1 0 2)。学習部 1 5 1 は、音響特徴量を第 1 中間表現モデル 1 0 に入力し、音響特徴量の中間表現 m を算出する (ステップ S 1 0 3)。

【 0 0 6 7 】

学習部 1 5 1 は、音響特徴量の中間表現 m を第 2 中間表現モデル 1 1 に入力し、サブバンド信号の中間表現 x_t を算出する (ステップ S 1 0 4)。学習部 1 5 1 は、確率分布生成モデル 1 2 を用いて確率分布 y_t 、 b を算出し、確率分布 y_t 、 b を基にサブバンド信号 s_t 、 b を算出する (ステップ S 1 0 5)。

【 0 0 6 8 】

一方、学習部 1 5 1 のサブバンド信号計算部 3 0 は、音声波形を基にして、サブバンド信号を計算する (ステップ S 1 0 6)。

10

【 0 0 6 9 】

学習部 1 5 1 の損失計算部 3 1 は、サブバンド信号 $s_t = 1 \sim T/B$ 、 $b = 1 \sim B$ と、確率分布 $y_t = 1 \sim T/B$ 、 $b = 1 \sim B$ とを基にして、損失値を計算する (ステップ S 1 0 7)。

【 0 0 7 0 】

学習部 1 5 1 のモデル学習部 3 2 は、誤差逆伝播に基づいて、第 1 中間表現モデル 1 0、第 2 中間表現モデル 1 1、確率分布生成モデル 1 2 のパラメータを学習する (ステップ S 1 0 8)。

【 0 0 7 1 】

学習部 1 5 1 は、学習を継続する場合には (ステップ S 1 0 9, Yes)、ステップ S 1 0 1 に移行する。一方、学習部 1 5 1 は、学習を継続しない場合には (ステップ S 1 0 9, No)、処理を終了する。

20

【 0 0 7 2 】

図 6 は、本実施例 1 に係る生成装置の生成時の処理手順を示すフローチャートである。図 6 に示すように、生成装置 1 0 0 の生成部 1 5 2 は、音響特徴量を取得する (ステップ S 2 0 1)。生成部 1 5 2 は、音響特徴量を第 1 中間表現モデル 1 0 に入力し、音響特徴量の中間表現 m を算出する (ステップ S 2 0 2)。

【 0 0 7 3 】

生成部 1 5 2 は、音響特徴量の中間表現 m を第 2 中間表現モデル 1 1 に入力し、サブバンド信号の中間表現 x_t を算出する (ステップ S 2 0 3)。生成部 1 5 2 は、確率分布生成モデル 1 2 を用いて確率分布 y_t 、 b を算出し、確率分布 y_t 、 b を基にサブバンド信号 s_t 、 b を算出する (ステップ S 2 0 4)。

30

【 0 0 7 4 】

生成部 1 5 2 の音声波形変換部 4 0 は、サブバンド信号 $s_t = 1 \sim T/B$ 、 $b = 1 \sim B$ を基にして、音声波形 $z_t = 1 \sim T$ を算出する (ステップ S 2 0 5)。生成部 1 5 2 は、音声波形 $z_t = 1 \sim T$ を出力する (ステップ S 2 0 6)。

【 0 0 7 5 】

生成部 1 5 2 は、処理を継続する場合には (ステップ S 2 0 7, Yes)、ステップ S 2 0 1 に移行する。生成部 1 5 2 は、処理を継続しない場合には (ステップ S 2 0 7, No)、処理を終了する。

【 0 0 7 6 】

40

次に、本実施例 1 に係る生成装置 1 0 0 の効果について説明する。生成装置 1 0 0 は、 B 個のサブバンド信号を 1 度の順伝搬で同時に N サンプル分生成することで、音響特徴量から音声波形を生成する処理を最大 $B \times N$ 倍に高速化することができる。

【 0 0 7 7 】

たとえば、確率分布生成モデル 1 2 - 1 - 1 $\sim$ 1 2 - 1 - B とサンプリング部 2 1 - 1 - 1 $\sim$ 2 1 - 1 - B との組、及び、確率分布生成モデル 1 2 - 2 - 1 $\sim$ 1 2 - 2 - B とサンプリング部 2 1 - 2 - 1 $\sim$ 2 1 - 2 - B との組によって、1 つの中間表現 $x_t = -1, 0$ から、 $B \times N$ ($N = 2$) 個に相当するサブバンド信号 ($s_t = 1$ 、 $b = 1 \sim B$ 、 $s_t = 2$ 、 $b = 1 \sim B$) が生成される。

【 0 0 7 8 】

50

また、音響特徴量から音声波形を生成する処理を高速化することで、ニューラルボコーダーの順伝播の回数を削減でき、学習および推論速度が向上する。また、ニューラルボコーダーを搭載可能なデバイスの範囲を広げることや、高サンプリング周波数でのリアルタイム動作に有効である。

【実施例 2】

【0079】

まず、本実施例 2 に係る生成装置の構成例について説明する。図 7 は、本実施例 2 に係る生成装置の構成を示す機能ブロック図である。図 7 に示すように、この生成装置 200 は、通信制御部 210 と、入力部 220 と、出力部 230 と、記憶部 240 と、制御部 250 とを有する。

10

【0080】

通信制御部 210、入力部 220、出力部 230 に関する説明は、実施例 1 で説明した通信制御部 110、入力部 120、出力部 130 に関する説明と同様である。

【0081】

記憶部 240 は、学習用音声波形データ 141 と、第 1 中間表現モデル 10 と、第 2 中間表現モデル 11 と、同時確率分布生成モデル 13 とを有する。記憶部 240 は、RAM、フラッシュメモリ等の半導体メモリ素子、または、ハードディスク、光ディスク等の記憶装置によって実現される。

【0082】

学習用音声波形データ 141、第 1 中間表現モデル 10、第 2 中間表現モデル 11 は、実施例 1 で説明した学習用音声波形データ 141、第 1 中間表現モデル 10、第 2 中間表現モデル 11 に関する説明と同様である。

20

【0083】

同時確率分布生成モデル 13 は、チャンネル $b = 1 \sim B$ のサブバンド信号を同時に推定するモデルである。同時確率分布生成モデル 13 より得た同時確率から各サブバンド信号に該当する次元を確率分布とする。

【0084】

ここで、実施例 1 で説明した N サンプルを同時生成する確率分布生成モデル 12 から出力される確率分布 $y_{t=\tau, b}$ は、式 (1) のように示される。

【0085】

30

【数 1】

$$P(s_{t=\tau, b} | s_{t=1, b=1, \dots, B}, \dots, s_{t=\tau-N, b=1, \dots, B}, m) \quad \dots (1)$$

【0086】

本実施例 2 の同時確率分布生成モデル 13 は、チャンネル $b = 1 \sim B$ のサブバンド信号を同時に推定するため、同時確率分布生成モデル 13 から出力される確率分布 $y_{t=\tau, b=1 \sim B}$ は、式 (2) のように示される。これにより、サブバンド間の関係性を考慮した確率分布を推定することができる。

【0087】

40

【数 2】

$$P(s_{t=\tau, b=1, \dots, B} | s_{t=1, b=1, \dots, B}, \dots, s_{t=\tau-N, b=1, \dots, B}, m) \quad \dots (2)$$

【0088】

図 7 において、制御部 250 は、学習部 251 と、生成部 252 とを有する。制御部 250 は、CPU 等に対応する。

【0089】

学習部 251 は、学習用音声波形データ 141 を用いて、第 1 中間表現モデル 10、第 2 中間表現モデル 11、同時確率分布生成モデル 13 の学習を実行する。

50

【 0 0 9 0 】

生成部 2 5 2 は、学習済みの第 1 中間表現モデル 1 0、第 2 中間表現モデル 1 1、同時確率分布生成モデル 1 3 を用いて、音響特徴量から音声波形を生成する。生成部 2 5 2 は、音響特徴量の情報を、入力部 2 2 0 から取得してもよいし、通信制御部 2 1 0 を介して、外部装置から取得してもよい。

【 0 0 9 1 】

次に、図 7 に示した学習部 2 5 1 について具体的に説明する。図 8 は、本実施例 2 に係る学習部を説明するための図である。図 8 に示すように、学習部 2 5 1 は、音響特徴量計算部 2 0、第 1 中間表現モデル 1 0、複数の第 2 中間表現モデル 1 1 - n ($n = 1 \sim T/B - 1$)、複数の同時確率分布生成モデル 1 3 - t ($t = 1、\dots、T/B$)、サンプリング部 2 1 - $t - b$ 、サブバンド信号計算部 3 0、損失計算部 3 1、モデル学習部 3 2 を有する。ここでは、音声の系列長を T 、サブバンド信号のチャンネル数を B 、サブバンド信号のサンプル数を $N (= 2)$ とする。

10

【 0 0 9 2 】

学習部 2 5 1 は、学習用音声波形データ 1 4 1 に格納された音声波形 $D 1$ を音響特徴量計算部 2 0 に入力する。音響特徴量計算部 2 0 は、音声波形 $D 1$ を基にして、音響特徴量 $D 2$ を算出する。音響特徴量計算部 2 0 は、音響特徴量 $D 2$ を、第 1 中間表現モデル 1 0 に入力する。

【 0 0 9 3 】

第 1 中間表現モデル 1 0 は、音響特徴量 $D 2$ が入力された場合に、音響特徴量の中間表現 m を出力する。第 1 中間表現モデル 1 0 は、中間表現 m を、第 2 中間表現モデル 1 1 - $1 \sim 1 1 - n$ に出力する。第 1 中間表現モデル 1 0 は、勾配情報を、モデル学習部 3 2 に出力する。

20

【 0 0 9 4 】

各第 2 中間表現モデル 1 1 - $1 \sim 1 1 - n$ は、中間表現 m が入力された場合に、サブバンド信号の中間表現 x_t を出力する。たとえば、第 2 中間表現モデル 1 1 - 1 は、中間表現 $x_{t=1,0}$ を算出し、同時確率分布生成モデル 1 3 - $1, 1 3 - 2$ に出力する。なお、第 2 中間表現モデル 1 1 - $2 \sim 1 1 - n$ は、サブバンド信号の中間表現 x_t を計算する場合に、前の時刻のサブバンド信号 $s_{t,b}$ を更に用いる。

【 0 0 9 5 】

第 2 中間表現モデル 1 1 - 2 は、中間表現 $x_{t=1,2}$ を算出し、同時確率分布生成モデル 1 2 - $3, 1 2 - 4$ に出力する。第 2 中間表現モデル 1 1 - 3 は、中間表現 $x_{t=3,4}$ を算出し、同時確率分布生成モデル 1 2 - $5, 1 2 - 6$ (図示略) に出力する。

30

【 0 0 9 6 】

以下の説明では、第 2 中間表現モデル 1 1 - $1 \sim 1 1 - n$ を区別しない場合、まとめて、第 2 中間表現モデル 1 1 と表記する。

【 0 0 9 7 】

各同時確率分布生成モデル 1 3 - t は、中間表現 x_t が入力された場合に、確率分布 $y_{t,b=1 \sim B}$ を出力する。たとえば、同時確率分布生成モデル 1 3 - 1 は、確率分布 $y_{t=1,b=1 \sim B}$ を、サンプリング部 2 1 - $1 - 1 \sim 2 1 - 1 - B$ に出力する。

40

【 0 0 9 8 】

同時確率分布生成モデル 1 3 - 1 が、チャンネル毎の確率分布を、それぞれ、サンプリング部 2 1 - $1 - 1 \sim 2 1 - 1 - B$ に出力してもよいし、サンプリング部 2 1 - $1 - 1 \sim 2 1 - 1 - B$ 側が、同時確率分布から、該当する確率分布をサンプリングして取得してもよい。たとえば、多変量分布を使えば、同時確率分布から該当する確率分布をサンプリングすることが可能である。以下に説明する他の同時確率分布生成モデルについても同様である。

【 0 0 9 9 】

同時確率分布生成モデル 1 3 - 2 は、確率分布 $y_{t=2,b=1 \sim B}$ を、サンプリング部 2 1 - $2 - 1 \sim 2 1 - 2 - B$ に出力する。

50

【 0 1 0 0 】

同時確率分布生成モデル 1 3 - 3 は、確率分布 $y_{t=3}$ 、 $b = 1 \sim B$ を、サンプリング部 2 1 - 3 - 1 ~ 2 1 - 3 - B に出力する。

【 0 1 0 1 】

同時確率分布生成モデル 1 3 - 4 は、確率分布 $y_{t=4}$ 、 $b = 1 \sim B$ を、サンプリング部 2 1 - 4 - 1 ~ 2 1 - 4 - B に出力する。

【 0 1 0 2 】

その他の同時確率分布生成モデルの説明を省略するが、それぞれ、中間表現 x_t が入力された場合に、確率分布 y_t 、 $b = 1 \sim B$ を、配下のサンプリング部に出力する。以下の説明では、各同時確率分布生成モデル 1 3 - t を区別しない場合、まとめて、同時確率分布生成モデル 1 3 と表記する。同時確率分布生成モデル 1 3 は、勾配情報を、モデル学習部 3 2 に出力する。同時確率分布生成モデル 1 3 は、確率分布 $y_{t=1 \sim T/B}$ 、 $b = 1 \sim B$ を、損失計算部 3 1 に出力する。

10

【 0 1 0 3 】

各サンプリング部 2 1 - t - b は、確率分布 y_t 、 b が入力された場合に、サブバンド信号 $s_{t,b}$ を算出する。なお、学習部 2 5 1 は、各同時確率分布生成モデル 1 3 が、直接サブバンド信号を予測する場合には、各サンプリング部 2 1 - t - b を用いなくてもよい。

【 0 1 0 4 】

各サンプリング部 2 1 - t - b の関するその他の説明は、実施例 1 で説明した各サンプリング部 2 1 - t - b に関する説明と同様である。

20

【 0 1 0 5 】

サブバンド信号計算部 3 0 は、音声波形 D 1 から、サブバンド信号 $s_{t=1 \sim T/B}$ 、 $b = 1 \sim B$ を計算する。サブバンド信号計算部 3 0 は、サブバンド信号 $s_{t=1 \sim T/B}$ 、 $b = 1 \sim B$ を、損失計算部 3 1 に出力する。サブバンド信号計算部 3 0 に関するその他の説明は、実施例 1 と同様である。

【 0 1 0 6 】

損失計算部 3 1 は、サブバンド信号 $s_{t=1 \sim T/B}$ 、 $b = 1 \sim B$ と、確率分布 $y_{t=1 \sim T/B}$ 、 $b = 1 \sim B$ を基にして、損失値 D 3 を算出する。損失計算部 3 1 は、損失値 D 3 を、モデル学習部 3 2 に出力する。損失計算部 3 1 に関するその他の説明は、実施例 1 と同様である。

30

【 0 1 0 7 】

モデル学習部 3 2 は、損失値 D 3、第 1 中間表現モデル 1 0 の勾配情報、第 2 中間表現モデル 1 1 の勾配情報、同時確率分布生成モデル 1 3 の勾配情報を取得し、損失値 D 3 が小さくなるように、第 1 中間表現モデル 1 0 のパラメータ、第 2 中間表現モデル 1 1 のパラメータ、同時確率分布生成モデル 1 3 のパラメータを機械学習する。たとえば、モデル学習部 3 2 は、誤差逆伝播アルゴリズムを利用して、機械学習を実行する。

【 0 1 0 8 】

学習部 2 5 1 に含まれる各処理部は、学習用音声波形データ 1 4 1 に含まれる複数の音声波形 D 1 について、上記処理を繰り返し実行することで、学習済みの第 1 中間表現モデル 1 0、第 2 中間表現モデル 1 1、同時確率分布生成モデル 1 3 を得る。

40

【 0 1 0 9 】

次に、図 7 に示した生成部 2 5 2 について具体的に説明する。図 9 は、本実施例 2 に係る生成部を説明するための図である。図 9 に示すように、生成部 2 5 2 は、第 1 中間表現モデル 1 0、複数の第 2 中間表現モデル 1 1 - n ($n = 1 \sim T/B - 1$)、複数の同時確率分布生成モデル 1 3 - t ($t = 1, \dots, T/B$)、サンプリング部 2 1 - t - b 、音声波形変換部 4 0 を有する。ここでは、音声の系列長を T 、サブバンド信号のチャンネル数を B 、サブバンド信号のサンプル数を $N (= 2)$ とする。

【 0 1 1 0 】

図 8 の場合と同様にして、第 2 中間表現モデル 1 1 - 1 ~ 1 1 - n を区別しない場合、まとめて、第 2 中間表現モデル 1 1 と表記する。同時確率分布生成モデル 1 3 - t を区別

50

しない場合、まとめて、同時確率分布生成モデル 1 3 と表記する。

【 0 1 1 1 】

第 1 中間表現モデル 1 0、第 2 中間表現モデル 1 1、同時確率分布生成モデル 1 3 は、学習部 2 5 1 によって学習済みとする。

【 0 1 1 2 】

生成部 2 5 2 は、音響特徴量 D_{10} を、第 1 中間表現モデル 1 0 に入力する。第 1 中間表現モデル 1 0 は、音響特徴量 D_{10} が入力された場合に、音響特徴量の中間表現 m を出力する。第 1 中間表現モデル 1 0 は、中間表現 m を、第 2 中間表現モデル 1 1 - 1 ~ 1 1 - n に出力する。

【 0 1 1 3 】

第 2 中間表現モデル 1 1 は、中間表現 m が入力された場合に、サブバンド信号の中間表現 x_t を、同時確率分布生成モデル 1 3 に出力する。第 2 中間表現モデル 1 1 に関する説明は、図 8 で行った、各第 2 中間表現モデル 1 1 - 1 ~ 1 1 - n に関する説明と同様である。

【 0 1 1 4 】

同時確率分布生成モデル 1 3 は、中間表現 x_t が入力された場合に、確率分布 $y_t, b = 1 \sim B$ を、各サンプリング部 2 1 - $t - b$ に出力する。同時確率分布生成モデル 1 3 に関する説明は、図 8 で行った、各同時確率分布生成モデル 1 2 - t に関する説明と同様である。

【 0 1 1 5 】

サンプリング部 2 1 - $t - b$ は、確率分布 y_t, b が入力された場合に、サブバンド信号 s_t, b を算出する。各サンプリング部 2 1 - $t - b$ は、サブバンド信号 $s_t = 1 \sim T/B, b = 1 \sim B$ を、音声波形変換部 4 0 に出力する。

【 0 1 1 6 】

音声波形変換部 4 0 は、サブバンド信号 $s_t = 1 \sim T/B, b = 1 \sim B$ を基にして、音声波形 $z_t = 1 \sim T$ を算出する。音声波形変換部 4 0 は、直交ミラーフィルタや、疑似直交ミラーフィルタの逆変換によって、サブバンド信号を音声波形に変換する。

【 0 1 1 7 】

次に、本実施例 2 に係る生成装置 2 0 0 の処理手順の一例について説明する。図 1 0 は、本実施例 2 に係る生成装置の学習時の処理手順を示すフローチャートである。図 1 0 に示すように、生成装置 2 0 0 の学習部 2 5 1 は、学習用音声波形データ 1 4 1 から、音声波形を取得する（ステップ S 3 0 1）。

【 0 1 1 8 】

学習部 2 5 1 の音響特徴量計算部 2 0 は、音声波形から音響特徴量を計算する（ステップ S 3 0 2）。学習部 2 5 1 は、音響特徴量を第 1 中間表現モデル 1 0 に入力し、音響特徴量の中間表現 m を算出する（ステップ S 3 0 3）。

【 0 1 1 9 】

学習部 2 5 1 は、音響特徴量の中間表現 m を第 2 中間表現モデル 1 1 に入力し、サブバンド信号の中間表現 x_t を算出する（ステップ S 3 0 4）。学習部 2 5 1 は、同時確率分布生成モデル 1 3 を用いて確率分布 y_t, b を算出し、確率分布 y_t, b を基にサブバンド信号 s_t, b を算出する（ステップ S 3 0 5）。

【 0 1 2 0 】

一方、学習部 2 5 1 のサブバンド信号計算部 3 0 は、音声波形を基にして、サブバンド信号を計算する（ステップ S 3 0 6）。

【 0 1 2 1 】

学習部 2 5 1 の損失計算部 3 1 は、サブバンド信号 $s_t = 1 \sim T/B, b = 1 \sim B$ と、確率分布 $y_t = 1 \sim T/B, b = 1 \sim B$ とを基にして、損失値を計算する（ステップ S 3 0 7）。

【 0 1 2 2 】

学習部 2 5 1 のモデル学習部 3 2 は、誤差逆伝播に基づいて、第 1 中間表現モデル 1 0、第 2 中間表現モデル 1 1、同時確率分布生成モデル 1 3 のパラメータを学習する（ステップ S 3 0 8）。

10

20

30

40

50

【 0 1 2 3 】

学習部 2 5 1 は、学習を継続する場合には（ステップ S 3 0 9 , Y e s ）、ステップ S 3 0 1 に移行する。一方、学習部 2 5 1 は、学習を継続しない場合には（ステップ S 3 0 9 , N o ）、処理を終了する。

【 0 1 2 4 】

図 1 1 は、本実施例 2 に係る生成装置の生成時の処理手順を示すフローチャートである。図 1 1 に示すように、生成装置 2 0 0 の生成部 2 5 2 は、音響特徴量を取得する（ステップ S 4 0 1 ）。生成部 2 5 2 は、音響特徴量を第 1 中間表現モデル 1 0 に入力し、音響特徴量の中間表現 m を算出する（ステップ S 4 0 2 ）。

【 0 1 2 5 】

生成部 2 5 2 は、音響特徴量の中間表現 m を第 2 中間表現モデル 1 1 に入力し、サブバンド信号の中間表現 x_t を算出する（ステップ S 4 0 3 ）。生成部 2 5 2 は、同時確率分布生成モデル 1 3 を用いて確率分布 y_t, b を算出し、確率分布 y_t, b を基にサブバンド信号 s_t, b を算出する（ステップ S 4 0 4 ）。

【 0 1 2 6 】

生成部 2 5 2 の音声波形変換部 4 0 は、サブバンド信号 $s_t = 1 \sim T/B, b = 1 \sim B$ を基にして、音声波形 $z_t = 1 \sim T$ を算出する（ステップ S 4 0 5 ）。生成部 2 5 2 は、音声波形 $z_t = 1 \sim T$ を出力する（ステップ S 4 0 6 ）。

【 0 1 2 7 】

生成部 2 5 2 は、処理を継続する場合には（ステップ S 4 0 7 , Y e s ）、ステップ S 4 0 1 に移行する。生成部 2 5 2 は、処理を継続しない場合には（ステップ S 4 0 7 , N o ）、処理を終了する。

【 0 1 2 8 】

次に、本実施例 2 に係る生成装置 2 0 0 の効果について説明する。生成装置 2 0 0 は、 B 個のサブバンド信号を 1 度の順伝搬で同時に N サンプル分生成することで、音響特徴量から音声波形を生成する処理を最大 $B \times N$ 倍に高速化することができる。また、生成装置 2 0 0 は、同時確率分布生成モデル 1 3 を利用することで、サブバンド間の関係性を考慮した確率分布を推定することができる。確率分布計算を複数バンド分同時に実行するため、行列演算命令をコールする回数を実施例 1 と比較して、 $1/B$ にすることができ、処理面でも効率化を図ることができる。

【実施例 3】

【 0 1 2 9 】

まず、本実施例 3 に係る生成装置の構成例について説明する。図 1 2 は、本実施例 3 に係る生成装置の構成を示す機能ブロック図である。図 1 2 に示すように、この生成装置 3 0 0 は、通信制御部 3 1 0 と、入力部 3 2 0 と、出力部 3 3 0 と、記憶部 3 4 0 と、制御部 3 5 0 とを有する。

【 0 1 3 0 】

通信制御部 3 1 0 、入力部 3 2 0 、出力部 3 3 0 に関する説明は、実施例 1 で説明した通信制御部 1 1 0 、入力部 1 2 0 、出力部 1 3 0 に関する説明と同様である。

【 0 1 3 1 】

記憶部 3 4 0 は、学習用音声波形データ 1 4 1 と、第 1 中間表現モデル 1 0 と、第 2 中間表現モデル 1 1 と、同時確率分布生成モデル 1 4 とを有する。記憶部 3 4 0 は、RAM、フラッシュメモリ等の半導体メモリ素子、または、ハードディスク、光ディスク等の記憶装置によって実現される。

【 0 1 3 2 】

学習用音声波形データ 1 4 1 、第 1 中間表現モデル 1 0 、第 2 中間表現モデル 1 1 は、実施例 1 で説明した学習用音声波形データ 1 4 1 、第 1 中間表現モデル 1 0 、第 2 中間表現モデル 1 1 に関する説明と同様である。

【 0 1 3 3 】

同時確率分布生成モデル 1 4 は、各時刻 $t = \tau, \tau + 1$ のチャンネル $b = 1 \sim B$ のサブ

10

20

30

40

50

バンド信号を同時に推定するモデルである。同時確率分布生成モデル 1 4 より得た同時確率から各サブバンド信号および各時刻に該当する次元を確率分布とする。

【 0 1 3 4 】

ここで、実施例 2 で説明した N サンプルを同時生成する同時確率分布 1 3 モデルから出力される確率分布は、それぞれ、式 (3) のように示される。たとえば、N = 2 の場合には、同時確率分布生成モデル 1 3 - 1 , 1 3 - 2 を用いて、各時刻 $t = \tau$, $\tau + 1$ のチャンネル $b = 1 \sim B$ のサブバンド信号を推定している。

【 0 1 3 5 】

【数 3 】

$$P(s_{t=\tau, b=1, \dots, B} | s_{t=1, b=1, \dots, B}, \dots, s_{t=\tau-N, b=1, \dots, B}, m), \dots, \dots (3)$$

$$P(s_{t=\tau+N-1, b=1, \dots, B} | s_{t=1, b=1, \dots, B}, \dots, s_{t=\tau-N, b=1, \dots, B}, m)$$

【 0 1 3 6 】

一方、本実施例 3 の同時確率分布生成モデル 1 4 は、各時刻、及び、チャンネル $b = 1 \sim B$ のサブバンド信号を同時に推定するため、同時確率分布生成モデル 1 4 から出力される確率分布は、式 (4) のように示される。これにより、複数のサブバンド間の関係性を考慮した確率分布を推定することができる。

【 0 1 3 7 】

【数 4 】

$$P(s_{t=\tau, b=1, \dots, B}, \dots, s_{t=\tau+N-1, b=1, \dots, B} | s_{t=1, b=1, \dots, B}, \dots, s_{t=\tau-N, b=1, \dots, B}, m) \dots (4)$$

【 0 1 3 8 】

図 1 2 において、制御部 3 5 0 は、学習部 3 5 1 と、生成部 3 5 2 とを有する。制御部 3 5 0 は、CPU 等に対応する。

【 0 1 3 9 】

学習部 3 5 1 は、学習用音声波形データ 1 4 1 を用いて、第 1 中間表現モデル 1 0 、第 2 中間表現モデル 1 1 、同時確率分布生成モデル 1 4 の学習を実行する。

【 0 1 4 0 】

生成部 3 5 2 は、学習済みの第 1 中間表現モデル 1 0 、第 2 中間表現モデル 1 1 、同時確率分布生成モデル 1 4 を用いて、音響特徴量から音声波形を生成する。生成部 3 5 2 は、音響特徴量の情報を、入力部 3 2 0 から取得してもよいし、通信制御部 3 1 0 を介して、外部装置から取得してもよい。

【 0 1 4 1 】

次に、図 1 2 に示した学習部 3 5 1 について具体的に説明する。図 1 3 は、本実施例 3 に係る学習部を説明するための図である。図 1 3 に示すように、学習部 3 5 1 は、音響特徴量計算部 2 0 、第 1 中間表現モデル 1 0 、複数の第 2 中間表現モデル 1 1 - n ($n = 1 \sim T / B - 1$) 、複数の同時確率分布生成モデル 1 4 - t ($t = 1, \dots, T / 2 B$) 、サンプリング部 2 1 - t - b 、サブバンド信号計算部 3 0 、損失計算部 3 1 、モデル学習部 3 2 を有する。ここでは、音声の系列長を T 、サブバンド信号のチャンネル数を B 、サブバンド信号のサンプル数を N (= 2) とする。

【 0 1 4 2 】

学習部 3 5 1 は、学習用音声波形データ 1 4 1 に格納された音声波形 D 1 を音響特徴量計算部 2 0 に入力する。音響特徴量計算部 2 0 は、音声波形 D 1 を基にして、音響特徴量 D 2 を算出する。音響特徴量計算部 2 0 は、音響特徴量 D 2 を、第 1 中間表現モデル 1 0 に入力する。

【 0 1 4 3 】

第 1 中間表現モデル 1 0 は、音響特徴量 D 2 が入力された場合に、音響特徴量の中間表

10

20

30

40

50

現 m を出力する。第 1 中間表現モデル 10 は、中間表現 m を、第 2 中間表現モデル 11 - 1 ~ 11 - n に出力する。第 1 中間表現モデル 10 は、勾配情報を、モデル学習部 32 に出力する。

【0144】

各第 2 中間表現モデル 11 - 1 ~ 11 - n は、中間表現 m が入力された場合に、サブバンド信号の中間表現 x_t を出力する。たとえば、第 2 中間表現モデル 11 - 1 は、中間表現 $x_t = -1, 0$ を算出し、同時確率分布生成モデル 14 - 1 に出力する。なお、第 2 中間表現モデル 11 - 2 ~ 11 - n は、サブバンド信号の中間表現 x_t を計算する場合に、前の時刻のサブバンド信号 s_t, b を更に用いる。

【0145】

第 2 中間表現モデル 11 - 2 は、中間表現 $x_t = 1, 2$ を算出し、同時確率分布生成モデル 14 - 2 に出力する。第 2 中間表現モデル 11 - 3 は、中間表現 $x_t = 3, 4$ を算出し、同時確率分布生成モデル 14 - 3 (図示略) に出力する。

【0146】

以下の説明では、第 2 中間表現モデル 11 - 1 ~ 11 - n を区別しない場合、まとめて、第 2 中間表現モデル 11 と表記する。

【0147】

各同時確率分布生成モデル 14 - t は、中間表現 x_t が入力された場合に、確率分布 $y_{t=t, t+1, b=1 \sim B}$ を出力する。たとえば、同時確率分布生成モデル 14 - 1 は、確率分布 $y_{t=1, b=1 \sim B}$ を、サンプリング部 21 - 1 - 1 ~ 21 - 1 - B に出力し、確率分布 $y_{t=2, b=1 \sim B}$ を、サンプリング部 21 - 2 - 1 ~ 21 - 2 - B に出力する。

【0148】

同時確率分布生成モデル 14 - 1 が、時刻毎、チャンネル毎の確率分布を、それぞれ、サンプリング部 21 - 1 - 1 ~ 21 - 1 - $B, 21 - 2 - 1 \sim 21 - 2 - B$ に出力してもよいし、サンプリング部 21 - 1 - 1 ~ 21 - 1 - $B, 21 - 2 - 1 \sim 21 - 2 - B$ 側が、同時確率分布から、該当する確率分布をサンプリングして取得してもよい。たとえば、多変量分布を使えば、同時確率分布から該当する確率分布をサンプリングすることが可能である。以下に説明する他の同時確率分布生成モデルについても同様である。

【0149】

同時確率分布生成モデル 14 - 2 は、確率分布 $y_{t=3, b=1 \sim B}$ を、サンプリング部 21 - 3 - 1 ~ 21 - 3 - B に出力し、確率分布 $y_{t=4, b=1 \sim B}$ を、サンプリング部 21 - 4 - 1 ~ 21 - 4 - B に出力する。

【0150】

その他の同時確率分布生成モデルの説明を省略するが、それぞれ、中間表現 x_t が入力された場合に、確率分布 $y_{t, t+1, b=1 \sim B}$ を、配下のサンプリング部に出力する。以下の説明では、各同時確率分布生成モデル 14 を区別しない場合、まとめて、同時確率分布生成モデル 14 と表記する。同時確率分布生成モデル 14 は、勾配情報を、モデル学習部 32 に出力する。同時確率分布生成モデル 14 は、確率分布 $y_{t=1 \sim T/B, b=1 \sim B}$ を、損失計算部 31 に出力する。

【0151】

各サンプリング部 21 - t - b は、確率分布 $y_{t, b}$ が入力された場合に、サブバンド信号 s_t, b を算出する。なお、学習部 351 は、各同時確率分布生成モデル 14 が、直接サブバンド信号を予測する場合には、各サンプリング部 21 - t - b を用いなくてもよい。

【0152】

各サンプリング部 21 - t - b の関するその他の説明は、実施例 1 で説明した各サンプリング部 21 - t - b に関する説明と同様である。

【0153】

サブバンド信号計算部 30 は、音声波形 D_1 から、サブバンド信号 $s_{t=1 \sim T/B, b=1 \sim B}$ を計算する。サブバンド信号計算部 30 は、サブバンド信号 $s_{t=1 \sim T/B, b=1 \sim B}$ を、損失計算部 31 に出力する。サブバンド信号計算部 30 に関するその他の説明は

10

20

30

40

50

、実施例 1 と同様である。

【 0 1 5 4 】

損失計算部 3 1 は、サブバンド信号 $s_{t=1 \sim T/B}$ 、 $b=1 \sim B$ と、確率分布 $y_{t=1 \sim T/B}$ 、 $b=1 \sim B$ とを基にして、損失値 D_3 を算出する。損失計算部 3 1 は、損失値 D_3 を、モデル学習部 3 2 に出力する。損失計算部 3 1 に関するその他の説明は、実施例 1 と同様である。

【 0 1 5 5 】

モデル学習部 3 2 は、損失値 D_3 、第 1 中間表現モデル 1 0 の勾配情報、第 2 中間表現モデル 1 1 の勾配情報、同時確率分布生成モデル 1 4 の勾配情報を取得し、損失値 D_3 が小さくなるように、第 1 中間表現モデル 1 0 のパラメータ、第 2 中間表現モデル 1 1 のパラメータ、同時確率分布生成モデル 1 4 のパラメータを機械学習する。たとえば、モデル学習部 3 2 は、誤差逆伝播アルゴリズムを利用して、機械学習を実行する。

10

【 0 1 5 6 】

学習部 3 5 1 に含まれる各処理部は、学習用音声波形データ 1 4 1 に含まれる複数の音声波形 D_1 について、上記処理を繰り返し実行することで、学習済みの第 1 中間表現モデル 1 0、第 2 中間表現モデル 1 1、同時確率分布生成モデル 1 4 を得る。

【 0 1 5 7 】

次に、図 1 2 に示した生成部 3 5 2 について具体的に説明する。図 1 4 は、本実施例 3 に係る生成部を説明するための図である。図 1 4 に示すように、生成部 3 5 2 は、第 1 中間表現モデル 1 0、複数の第 2 中間表現モデル 1 1 - n ($n=1 \sim T/B-1$)、複数の同時確率分布生成モデル 1 4 - t ($t=1, \dots, T/2B$)、サンプリング部 2 1 - $t-b$ 、音声波形変換部 4 0 を有する。ここでは、音声の系列長を T 、サブバンド信号のチャンネル数を B 、サブバンド信号のサンプル数を $N (=2)$ とする。

20

【 0 1 5 8 】

図 1 3 の場合と同様にして、第 2 中間表現モデル 1 1 - $1 \sim 1 1 - n$ を区別しない場合、まとめて、第 2 中間表現モデル 1 1 と表記する。同時確率分布生成モデル 1 4 - t を区別しない場合、まとめて、同時確率分布生成モデル 1 4 と表記する。

【 0 1 5 9 】

第 1 中間表現モデル 1 0、第 2 中間表現モデル 1 1、同時確率分布生成モデル 1 4 は、学習部 3 5 1 によって学習済みとする。

30

【 0 1 6 0 】

生成部 3 5 2 は、音響特徴量 D_{10} を、第 1 中間表現モデル 1 0 に入力する。第 1 中間表現モデル 1 0 は、音響特徴量 D_{10} が入力された場合に、音響特徴量の中間表現 m を出力する。第 1 中間表現モデル 1 0 は、中間表現 m を、第 2 中間表現モデル 1 1 - $1 \sim 1 1 - n$ に出力する。

【 0 1 6 1 】

第 2 中間表現モデル 1 1 は、中間表現 m が入力された場合に、サブバンド信号の中間表現 x_t を、同時確率分布生成モデル 1 4 に出力する。第 2 中間表現モデル 1 1 に関する説明は、図 1 3 で行った、各第 2 中間表現モデル 1 1 - $1 \sim 1 1 - n$ に関する説明と同様である。

40

【 0 1 6 2 】

同時確率分布生成モデル 1 4 は、中間表現 x_t が入力された場合に、確率分布 $y_{t, t+1}$ 、 $b=1 \sim B$ を、各サンプリング部 2 1 - $t-b$ に出力する。同時確率分布生成モデル 1 4 に関する説明は、図 1 3 で行った、各同時確率分布生成モデル 1 2 - t に関する説明と同様である。

【 0 1 6 3 】

サンプリング部 2 1 - $t-b$ は、確率分布 $y_{t, b}$ が入力された場合に、サブバンド信号 $s_{t, b}$ を算出する。各サンプリング部 2 1 - $t-b$ は、サブバンド信号 $s_{t=1 \sim T/B}$ 、 $b=1 \sim B$ を、音声波形変換部 4 0 に出力する。

【 0 1 6 4 】

50

音声波形変換部 40 は、サブバンド信号 $s_{t,b} = 1 \sim T/B, b = 1 \sim B$ を基にして、音声波形 $z_t = 1 \sim T$ を算出する。音声波形変換部 40 は、直交ミラーフィルタや、疑似直交ミラーフィルタの逆変換によって、サブバンド信号を音声波形に変換する。

【0165】

次に、本実施例 3 に係る生成装置 300 の処理手順の一例について説明する。図 15 は、本実施例 3 に係る生成装置の学習時の処理手順を示すフローチャートである。図 15 に示すように、生成装置 300 の学習部 351 は、学習用音声波形データ 141 から、音声波形を取得する（ステップ S501）。

【0166】

学習部 351 の音響特徴量計算部 20 は、音声波形から音響特徴量を計算する（ステップ S502）。学習部 351 は、音響特徴量を第 1 中間表現モデル 10 に入力し、音響特徴量の中間表現 m を算出する（ステップ S503）。

10

【0167】

学習部 351 は、音響特徴量の中間表現 m を第 2 中間表現モデル 11 に入力し、サブバンド信号の中間表現 x_t を算出する（ステップ S504）。学習部 351 は、同時確率分布生成モデル 14 を用いて確率分布 $y_{t,t+1,b}$ を算出し、確率分布 $y_{t,t+1,b}$ を基にサブバンド信号 $s_{t,b}$ を算出する（ステップ S505）。

【0168】

一方、学習部 351 のサブバンド信号計算部 30 は、音声波形を基にして、サブバンド信号を計算する（ステップ S506）。

20

【0169】

学習部 351 の損失計算部 31 は、サブバンド信号 $s_{t,b} = 1 \sim T/B, b = 1 \sim B$ と、確率分布 $y_{t,b} = 1 \sim T/B, b = 1 \sim B$ とを基にして、損失値を計算する（ステップ S507）。

【0170】

学習部 351 のモデル学習部 32 は、誤差逆伝播に基づいて、第 1 中間表現モデル 10、第 2 中間表現モデル 11、同時確率分布生成モデル 14 のパラメータを学習する（ステップ S508）。

【0171】

学習部 351 は、学習を継続する場合には（ステップ S509, Yes）、ステップ S501 に移行する。一方、学習部 351 は、学習を継続しない場合には（ステップ S509, No）、処理を終了する。

30

【0172】

図 16 は、本実施例 3 に係る生成装置の生成時の処理手順を示すフローチャートである。図 16 に示すように、生成装置 300 の生成部 352 は、音響特徴量を取得する（ステップ S601）。生成部 352 は、音響特徴量を第 1 中間表現モデル 10 に入力し、音響特徴量の中間表現 m を算出する（ステップ S602）。

【0173】

生成部 352 は、音響特徴量の中間表現 m を第 2 中間表現モデル 11 に入力し、サブバンド信号の中間表現 x_t を算出する（ステップ S603）。生成部 352 は、同時確率分布生成モデル 15 を用いて確率分布 $y_{t,t+1,b}$ を算出し、確率分布 $y_{t,t+1,b}$ を基にサブバンド信号 $s_{t,b}$ を算出する（ステップ S604）。

40

【0174】

生成部 352 の音声波形変換部 40 は、サブバンド信号 $s_{t,b} = 1 \sim T/B, b = 1 \sim B$ を基にして、音声波形 $z_t = 1 \sim T$ を算出する（ステップ S605）。生成部 352 は、音声波形 $z_t = 1 \sim T$ を出力する（ステップ S606）。

【0175】

生成部 352 は、処理を継続する場合には（ステップ S607, Yes）、ステップ S601 に移行する。生成部 352 は、処理を継続しない場合には（ステップ S607, No）、処理を終了する。

【0176】

50

次に、本実施例 3 に係る生成装置 300 の効果について説明する。生成装置 300 は、各時刻に関する B 個のサブバンド信号を 1 度の順伝搬で同時に N サンプル分生成する。これによって、音声特徴量から音声波形を生成する処理を高速化できるとともに、行列演算命令をコールする回数を実施例 1 と比較して、最小で $1 / (B \times N)$ 回にすることができ、処理面でも有効である。

【0177】

続いて、生成プログラムを実行するコンピュータの一例について説明する。図 17 は、生成プログラムを実行するコンピュータの一例を示す図である。コンピュータ 1000 は、たとえば、メモリ 1010 と、CPU 1020 と、ハードディスクドライブインタフェース 1030 と、ディスクドライブインタフェース 1040 と、シリアルポートインタフェース 1050 と、ビデオアダプタ 1060 と、ネットワークインタフェース 1070 とを有する。これらの各部は、バス 1080 によって接続される。

10

【0178】

メモリ 1010 は、ROM (Read Only Memory) 1011 および RAM 1012 を含む。ROM 1011 は、たとえば、BIOS (Basic Input Output System) 等のブートプログラムを記憶する。ハードディスクドライブインタフェース 1030 は、ハードディスクドライブ 1031 に接続される。ディスクドライブインタフェース 1040 は、ディスクドライブ 1041 に接続される。ディスクドライブ 1041 には、たとえば、磁気ディスクや光ディスク等の着脱可能な記憶媒体が挿入される。シリアルポートインタフェース 1050 には、たとえば、マウス 1051 およびキーボード 1052 が接続される。ビデオアダプタ 1060 には、たとえば、ディスプレイ 1061 が接続される。

20

【0179】

ここで、ハードディスクドライブ 1031 は、たとえば、OS 1091、アプリケーションプログラム 1092、プログラムモジュール 1093 およびプログラムデータ 1094 を記憶する。上記実施形態で説明した各情報は、たとえばハードディスクドライブ 1031 やメモリ 1010 に記憶される。

【0180】

また、生成プログラムは、たとえば、コンピュータ 1000 によって実行される指令が記述されたプログラムモジュール 1093 として、ハードディスクドライブ 1031 に記憶される。具体的には、上記実施形態で説明した生成装置 100 (200, 300) が実行する各処理が記述されたプログラムモジュール 1093 が、ハードディスクドライブ 1031 に記憶される。

30

【0181】

また、生成プログラムによる情報処理に用いられるデータは、プログラムデータ 1094 として、たとえば、ハードディスクドライブ 1031 に記憶される。そして、CPU 1020 が、ハードディスクドライブ 1031 に記憶されたプログラムモジュール 1093 やプログラムデータ 1094 を必要に応じて RAM 1012 に読み出して、上述した各手順を実行する。

【0182】

なお、生成プログラムに係るプログラムモジュール 1093 やプログラムデータ 1094 は、ハードディスクドライブ 1031 に記憶される場合に限られず、たとえば、着脱可能な記憶媒体に記憶されて、ディスクドライブ 1041 等を介して CPU 1020 によって読み出されてもよい。あるいは、生成プログラムに係るプログラムモジュール 1093 やプログラムデータ 1094 は、LAN や WAN (Wide Area Network) 等のネットワークを介して接続された他のコンピュータに記憶され、ネットワークインタフェース 1070 を介して CPU 1020 によって読み出されてもよい。

40

【0183】

以上、本発明者によってなされた発明を適用した実施形態について説明したが、本実施形態による本発明の開示の一部をなす記述および図面により本発明は限定されることはない。すなわち、本実施形態に基づいて当業者等によりなされる他の実施形態、実施例およ

50

び運用技術等は全て本発明の範疇に含まれる。

【符号の説明】

【 0 1 8 4 】

1 0 0 , 2 0 0 , 3 0 0	生成装置
1 1 0 , 2 1 0 , 3 1 0	通信制御部
1 2 0 , 2 2 0 , 3 2 0	入力部
1 3 0 , 2 3 0 , 3 3 0	出力部
1 4 0 , 2 4 0 , 3 4 0	記憶部
1 5 0 , 2 5 0 , 3 5 0	制御部
1 5 1 , 2 5 1 , 3 5 1	学習部
1 5 2 , 2 5 2 , 3 5 2	生成部

10

20

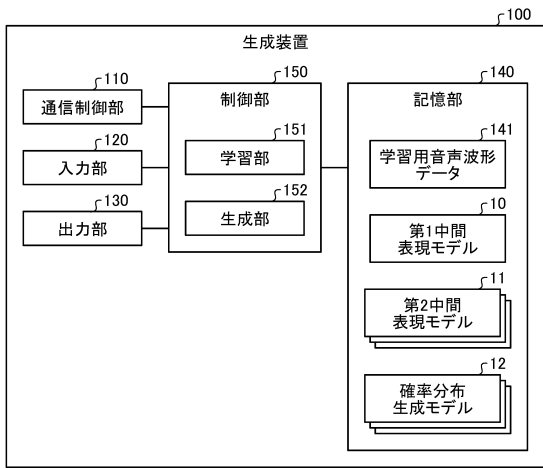
30

40

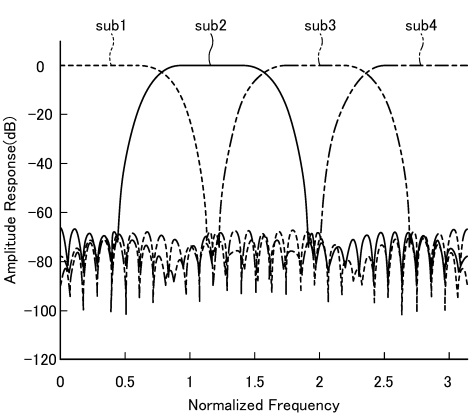
50

【図面】

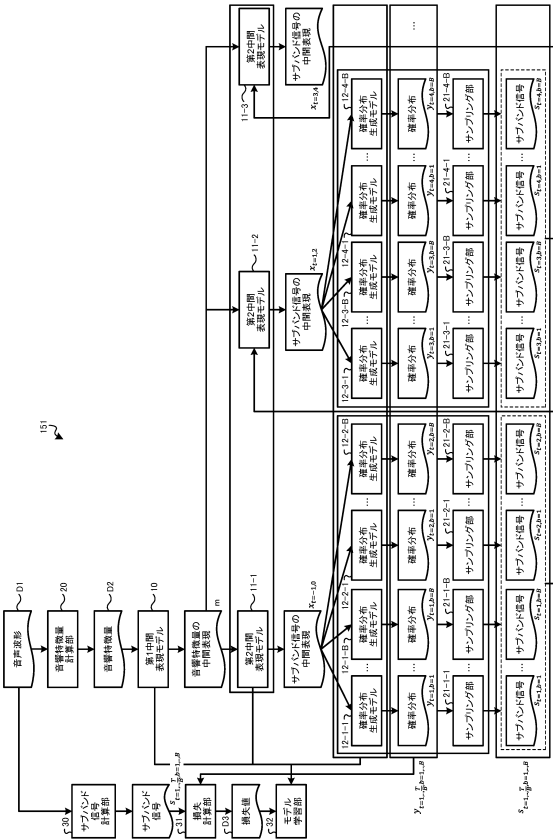
【図 1】



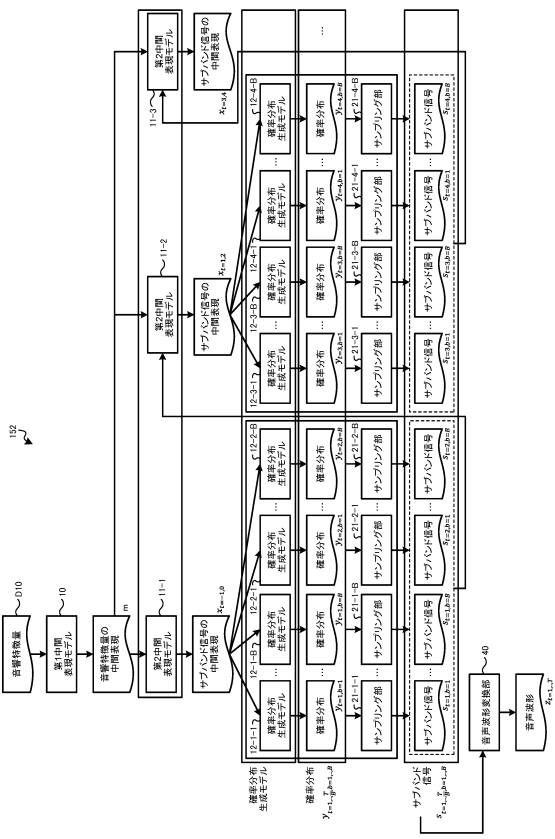
【図 2】



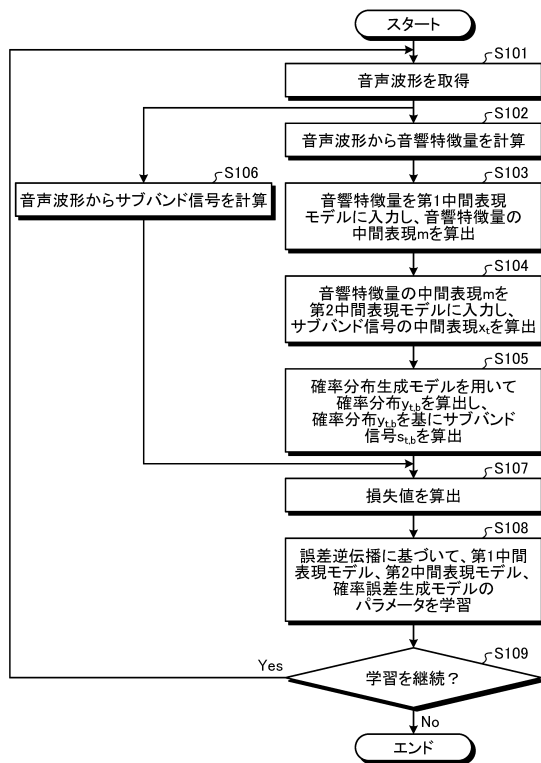
【図 3】



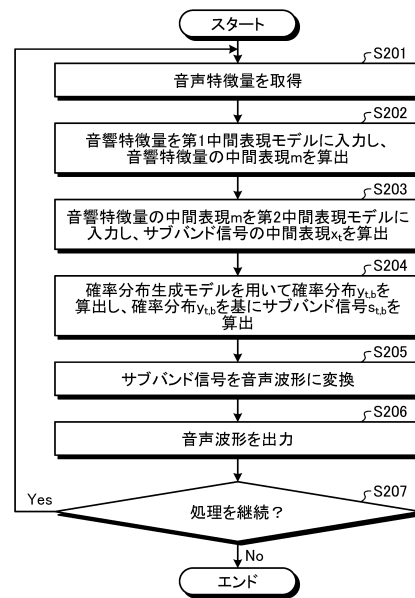
【図 4】



【 図 5 】



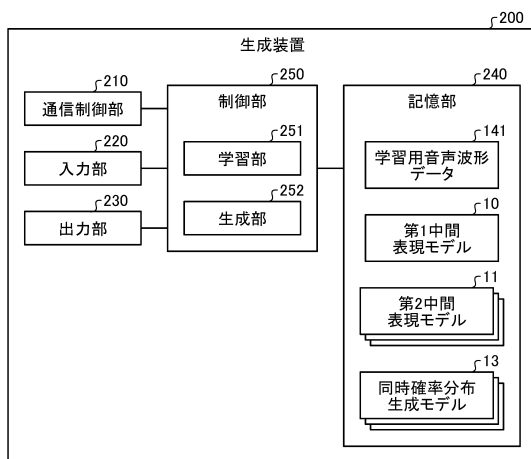
【 図 6 】



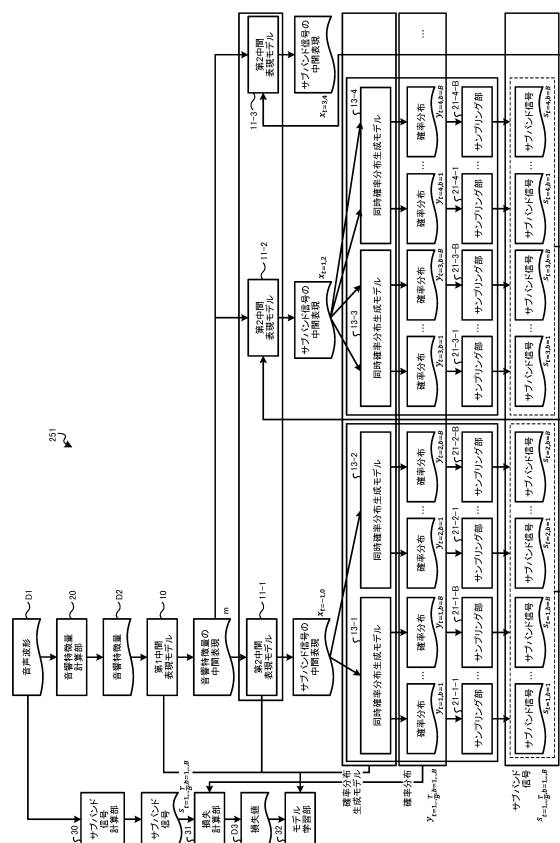
10

20

【圖 7】



【圖 8】

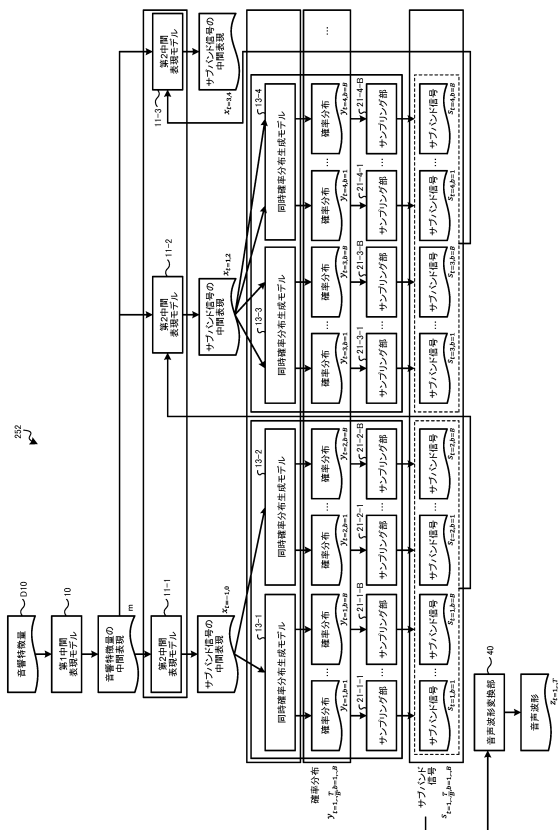


30

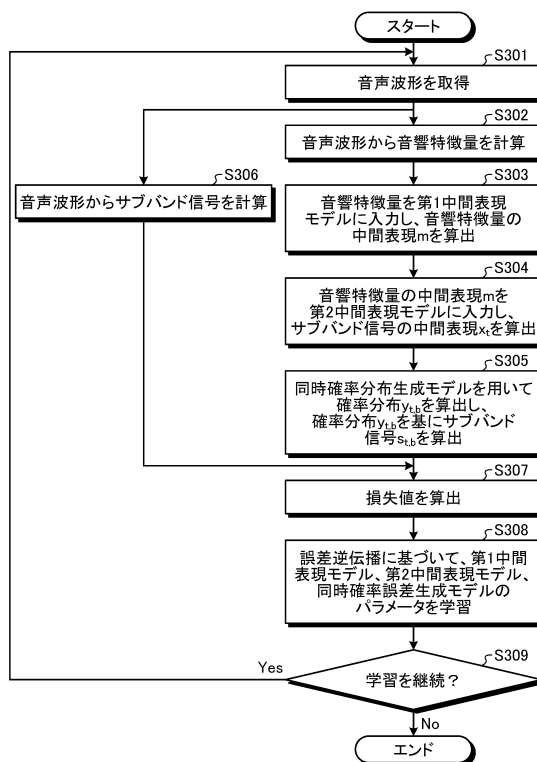
40

50

【図 9】



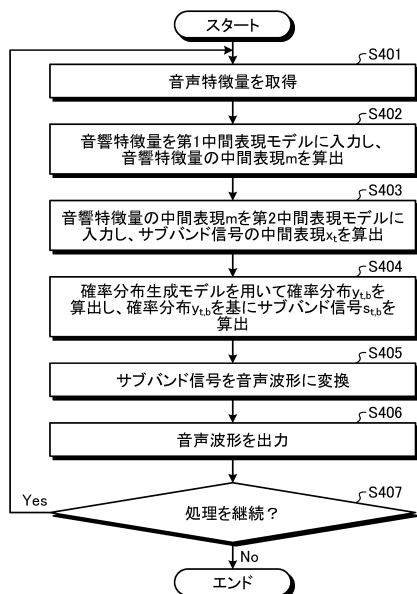
【図 10】



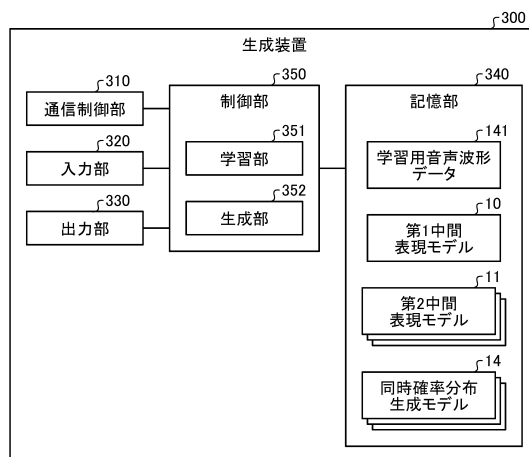
10

20

【図 11】



【図 12】

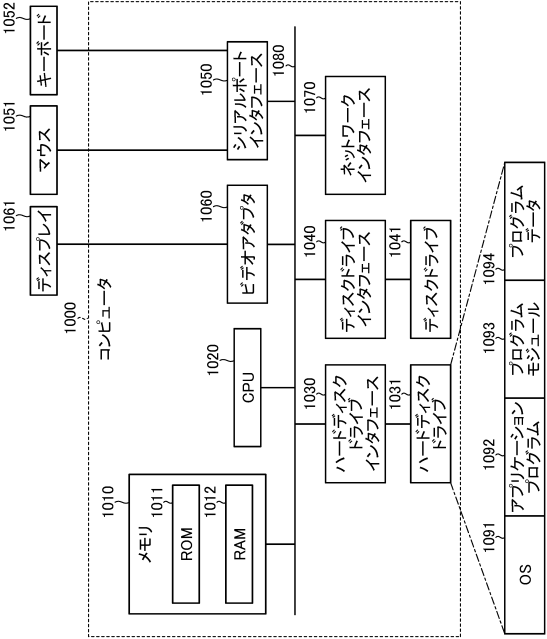


30

40

50

【図 17】



10

20

30

40

50

フロントページの続き

(56)参考文献 特開 2 0 2 1 - 3 2 9 3 7 (J P , A)
 特開 2 0 1 9 - 4 5 8 5 6 (J P , A)
 特開 2 0 1 8 - 1 4 1 9 1 7 (J P , A)
 特開 2 0 1 8 - 1 4 1 9 1 5 (J P , A)
 特開 2 0 2 1 - 6 7 8 8 5 (J P , A)
 国際公開第 2 0 2 0 / 1 4 5 4 7 2 (W O , A 1)
 岡本拓磨他，サブバンド WaveNet ボコーダによる全可聴帯域音声合成の検討，日本
 音響学会 2 0 1 8 年春季研究発表会講演論文集 [C D - R O M] ，2018年03月，pp.251-2
 54
(58)調査した分野 (Int.Cl. , D B 名)
 G 1 0 L 1 3 / 0 0 - 9 9 / 0 0