

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6078441号
(P6078441)

(45) 発行日 平成29年2月8日(2017.2.8)

(24) 登録日 平成29年1月20日(2017.1.20)

(51) Int.Cl.	F 1
G 10 L 15/10	(2006.01)
G 06 F 19/00	(2011.01)
G 10 L 15/06	(2013.01)
G 06 N 5/04	(2006.01)
	G 10 L 15/10
	G 06 F 19/00
	G 10 L 15/06
	G 06 N 5/04
	G 06 N 5/04

請求項の数 24 (全 71 頁)

(21) 出願番号	特願2013-182435 (P2013-182435)
(22) 出願日	平成25年9月3日(2013.9.3)
(65) 公開番号	特開2015-28579 (P2015-28579A)
(43) 公開日	平成27年2月12日(2015.2.12)
審査請求日	平成27年8月11日(2015.8.11)
(31) 優先権主張番号	特願2013-140029 (P2013-140029)
(32) 優先日	平成25年7月3日(2013.7.3)
(33) 優先権主張国	日本国(JP)

(73) 特許権者	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(74) 代理人	100121706 弁理士 中尾 直樹
(74) 代理人	100128705 弁理士 中村 幸雄
(74) 代理人	100147773 弁理士 義村 宗洋
(72) 発明者	井本 桂右 東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
(72) 発明者	植松 尚 東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】モデル処理装置、分析装置、それらの方法およびプログラム

(57) 【特許請求の範囲】

【請求項1】

少なくとも、各時間区間の音響信号を生じさせた音響イベントの時系列の並びを表す音響イベントラベル列、ならびに、前記音響イベントを生じさせた行動の候補を表す行動ラベル候補および前記音響イベントを生じさせた状況の候補の総数の少なくとも一方を用い、

前記音響イベントの時系列の並びにおける、前記行動の候補に対応する前記音響イベントの組み合わせと前記行動の候補の時系列の並びとの同時事後確率、

前記音響イベントの時系列の並びにおける、前記行動の候補に対応する前記音響イベントの組み合わせと前記行動の候補の時系列の並びと前記行動の候補に対応する前記状況の候補の組み合わせとの同時事後確率、または

前記音響イベントの時系列の並びにおける、前記状況の候補に対応する前記音響イベントの組み合わせと前記状況の候補の時系列の並びとの同時事後確率の何れかを最大化するための処理を含む学習処理によって、行動の候補が音響イベントを生成する確率 P (音響イベント | 行動の候補) および状況の候補が音響イベントを生成する確率 P (音響イベント | 状況の候補) の少なくとも一方を得るモデル化部を有するモデル処理装置。

【請求項2】

請求項1のモデル処理装置であって、

前記モデル化部は、

前記学習処理によって、行動の候補の時系列の遷移確率、および状況の候補の時系列の遷移確率の少なくとも一方を得る、モデル処理装置。

【請求項 3】

請求項 1 または 2 のモデル処理装置であって、

前記モデル化部は、

少なくとも、前記音響イベントラベル列および前記行動ラベル候補を用い、前記音響イベントの時系列の並びにおける、前記行動の候補に対応する前記音響イベントの組み合わせと前記行動の候補の時系列の並びとの同時事後確率を最大化するための処理を含む学習処理によって、少なくとも、前記確率 P (音響イベント | 行動の候補) を得る、モデル処理装置。

10

【請求項 4】

請求項 1 または 2 のモデル処理装置であって、

前記モデル化部は、

少なくとも、前記音響イベントラベル列、前記行動ラベル候補、および前記状況の候補の総数を用い、前記音響イベントの時系列の並びにおける、前記行動の候補に対応する前記音響イベントの組み合わせと前記行動の候補の時系列の並びと前記行動の候補に対応する前記状況の候補の組み合わせとの同時事後確率を最大化するための処理を含む学習処理によって、少なくとも、前記確率 P (音響イベント | 状況の候補)、および行動の候補が状況の候補を生成する確率 P (状況の候補 | 行動の候補) を得る、モデル処理装置。

【請求項 5】

請求項 1 または 2 のモデル処理装置であって、

前記モデル化部は、

少なくとも、前記音響イベントラベル列、前記行動ラベル候補、および前記状況の候補の総数を用い、前記音響イベントの時系列の並びにおける、前記行動の候補に対応する前記音響イベントの組み合わせと前記行動の候補の時系列の並びおよび/または状況の時系列の並びと前記行動の候補に対応する前記状況の候補の組み合わせとの同時事後確率を最大化するための処理を含む学習処理によって、

前記確率 P (音響イベント | 行動の候補) および前記確率 P (音響イベント | 状況の候補) の少なくとも一方を得る、モデル処理装置。

20

【請求項 6】

請求項 1 または 2 のモデル処理装置であって、

前記モデル化部は、

少なくとも、前記音響イベントラベル列、および前記状況の候補の総数を用い、前記音響イベントの時系列の並びにおける、前記状況の候補に対応する前記音響イベントの組み合わせと前記状況の候補の時系列の並びとの同時事後確率を最大化するための処理を含む学習処理によって、少なくとも、前記確率 P (音響イベント | 状況の候補) を得る、モデル処理装置。

30

【請求項 7】

請求項 1 から 6 の何れかのモデル処理装置であって、

前記音響信号に対応する音響特徴量を用いて前記音響イベントを決定し、前記音響イベントラベル列を得る音響イベント判定部を有する、モデル処理装置。

40

【請求項 8】

請求項 7 のモデル処理装置であって、

前記音響信号から音響特徴量を得る特徴量算出部を有し、

前記音響イベント判定部は、前記特徴量算出部で得られた前記音響特徴量を用いて前記音響イベントを決定し、

前記音響特徴量は、音圧レベル、音響パワー、M F C C 特徴量、L P C 特徴量、立ち上がり特性、調波性、時間周期性の何れかを含む、モデル処理装置。

【請求項 9】

少なくとも、行動の候補が音響イベントを生成する確率 P (音響イベント | 行動の

50

候補) および状況の候補が音響イベントを生成する確率 P (音響イベント | 状況の候補) の少なくとも一方、ならびに、各時間区間の音響信号を生じさせた第2音響イベントの時系列の並びを表す第2音響イベントラベル列を用い、

前記第2音響イベントの時系列の並びに対する、前記確率 P (音響イベント | 行動の候補) のもとでの、行動の候補の尤もらしさ、

前記第2音響イベントの分布と、行動の候補に対応する前記確率 P (音響イベント | 行動の候補) で表される音響イベントの分布との近さ、

前記第2音響イベントの時系列の並びに対する、前記確率 P (音響イベント | 状況の候補) のもとでの、状況の候補の尤もらしさ、および

前記第2音響イベントの分布と、状況の候補に対応する前記確率 P (音響イベント | 状況の候補) で表される音響イベントの分布との近さ、

の少なくとも何れかに基づいて、行動の候補の何れかおよび状況の候補の何れかの少なくとも一方を選択する生成モデル比較部

を有するモデル処理装置。

【請求項10】

少なくとも、請求項3の確率 P (音響イベント | 行動の候補)、ならびに

各時間区間の音響信号を生じさせた第2音響イベントの時系列の並びを表す第2音響イベントラベル列を用い、

前記第2音響イベントの時系列の並びに対する、前記確率 P (音響イベント | 行動の候補) のもとでの、行動の候補の尤もらしさ、および

前記第2音響イベントの分布と、行動の候補に対応する前記確率 P (音響イベント | 行動の候補) で表される音響イベントの分布との近さ、

の少なくとも何れかに基づいて、少なくとも、行動の候補の何れかを選択する生成モデル比較部を有するモデル処理装置。

【請求項11】

少なくとも、請求項4の確率 P (音響イベント | 行動の候補)、および前記確率 P (音響イベント | 状況の候補)、ならびに

各時間区間の音響信号を生じさせた第2音響イベントの時系列の並びを表す第2音響イベントラベル列を用い、

前記第2音響イベントの時系列の並びに対する、前記確率 P (音響イベント | 行動の候補) のもとでの、行動の候補の尤もらしさ、

前記第2音響イベントの分布と、行動の候補に対応する前記確率 P (音響イベント | 行動の候補) で表される音響イベントの分布との近さ、

前記第2音響イベントの時系列の並びに対する、前記確率 P (音響イベント | 状況の候補) のもとでの、状況の候補の尤もらしさ、および

前記第2音響イベントの分布と、状況の候補に対応する前記確率 P (音響イベント | 状況の候補) で表される音響イベントの分布との近さ、

の少なくとも何れかに基づいて、少なくとも、行動の候補の何れかおよび状況の候補の何れかを選択する生成モデル比較部を有するモデル処理装置。

【請求項12】

少なくとも、請求項5の確率 P (音響イベント | 状況の候補)、ならびに

各時間区間の音響信号を生じさせた第2音響イベントの時系列の並びを表す第2音響イベントラベル列を用い、

前記第2音響イベントの時系列の並びに対する、前記確率 P (音響イベント | 状況の候補) のもとでの、状況の候補の尤もらしさ、および

前記第2音響イベントの分布と、状況の候補に対応する前記確率 P (音響イベント | 状況の候補) で表される音響イベントの分布との近さ、

の少なくとも何れかに基づいて、少なくとも、状況の候補の何れかを選択する生成モデル比較部を有するモデル処理装置。

【請求項13】

10

20

30

40

50

請求項 1 から 3 の何れかのモデル処理装置であって、
 少なくとも、前記確率 P (音響イベント | 行動の候補)、ならびに
 各時間区間の音響信号を生じさせた第 2 音響イベントの時系列の並びを表す第 2 音
 韻イベントラベル列を用い、

前記第 2 音響イベントの時系列の並びに対する、前記確率 P (音響イベント | 行動
 の候補) のもとでの、行動の候補の尤もらしさ、および

前記第 2 音響イベントの分布と、行動の候補に対応する前記確率 P (音響イベント
 | 行動の候補) で表される音響イベントの分布との近さ、
 の少なくとも何れかに基づいて、行動の候補の何れかを選択する生成モデル比較部と、

少なくとも、前記前記第 2 音響イベントラベル列を含む第 3 音響イベントラベル列
 、および前記生成モデル比較部で選択された行動の候補を表す第 2 行動ラベル候補を含む
 第 3 行動ラベル候補を用い、

前記第 3 音響イベントラベル列が表す第 3 音響イベントの時系列の並びにおける、
 前記第 3 行動ラベル候補が表す行動の候補に対応する音響イベントの組み合わせと前記第
 3 行動ラベル候補が表す行動の候補の時系列の並びとの同時事後確率を最大化するための
 処理を含む学習処理によって、行動の候補が音響イベントを生成する確率 P 2 (音響イベ
 ント | 行動の候補) を得る第 2 モデル化部と、
 を有するモデル処理装置。

【請求項 1 4】

請求項 1 , 2 または 4 のモデル処理装置であって、
 少なくとも、前記確率 P (音響イベント | 行動の候補)、および前記確率 P (音響
 イベント | 状況の候補)、ならびに

各時間区間の音響信号を生じさせた第 2 音響イベントの時系列の並びを表す第 2 音
 韵イベントラベル列を用い、

前記第 2 音響イベントの時系列の並びに対する、前記確率 P (音響イベント | 行動
 の候補) のもとでの、行動の候補の尤もらしさ、

前記第 2 音響イベントの分布と、行動の候補に対応する前記確率 P (音響イベント
 | 行動の候補) で表される音響イベントの分布との近さ、

前記第 2 音響イベントの時系列の並びに対する、前記確率 P (音響イベント | 状況
 の候補) のもとでの、状況の候補の尤もらしさ、および

前記第 2 音響イベントの分布と、状況の候補に対応する前記確率 P (音響イベ
 ント | 状況の候補) で表される音響イベントの分布との近さ、
 の少なくとも何れかに基づいて、行動の候補の何れかおよび状況の候補の何れかを選択す
 る生成モデル比較部と、

少なくとも、前記第 2 音響イベントラベル列を含む第 3 音響イベントラベル列、前
 記生成モデル比較部で選択された行動の候補を表す第 2 行動ラベル候補を含む第 3 行動ラ
 ベル候補を用い、

前記第 3 音響イベントラベル列が表す第 3 音響イベントの時系列の並びにおける、
 前記第 3 行動ラベル候補が表す行動の候補に対応する前記第 3 音響イベントの組み合わせ
 と、前記第 3 行動ラベル候補が表す行動の候補の時系列の並びと、前記第 3 行動ラベル候
 补が表す行動の候補に対応する状況の候補の組み合わせと、の同時事後確率を最大化する
 ための処理を含む学習処理によって、行動の候補が音響イベントを生成する確率 P 2 (音
 韵イベント | 行動の候補)、および状況の候補が音響イベントを生成する確率 P 2 (音響
 イベント | 状況の候補) を得る第 2 モデル化部と、
 を有するモデル処理装置。

【請求項 1 5】

請求項 1 , 2 または 5 のモデル処理装置であって、
 少なくとも、前記確率 P (音響イベント | 状況の候補)、ならびに

各時間区間の音響信号を生じさせた第 2 音響イベントの時系列の並びを表す第 2 音
 韵イベントラベル列を用い、

10

20

30

40

50

前記第2音響イベントの時系列の並びに対する、前記確率P（音響イベント|状況の候補）のもとでの、状況の候補の尤もらしさ、および

前記第2音響イベントの分布と、状況の候補に対応する前記確率P（音響イベント|状況の候補）で表される音響イベントの分布との近さ、

の少なくとも何れかに基づいて、状況の候補の何れかを選択する生成モデル比較部と、

少なくとも、前記第2音響イベントラベル列を含む第3音響イベントラベル列を用い、

前記第3音響イベントラベル列が表す第3音響イベントの時系列の並びにおける、状況の候補に対応する前記第3音響イベントラベル列が表す第3音響イベントの組み合わせと当該状況の候補の時系列の並びとの同時事後確率を最大化するための処理を含む学習処理によって、状況の候補が音響イベントを生成する確率P2（音響イベント|状況の候補）を得る第2モデル化部と、

を有するモデル処理装置。

【請求項16】

請求項1から3の何れかのモデル処理装置であって

前記音響イベントの時系列の並びは、前記行動ラベル候補によって行動の候補が表された第4音響イベントの時系列の並び、および前記行動ラベル候補によって行動の候補が表されていない第5音響イベントの時系列の並びを含み、

前記モデル化部は、前記第5音響イベントを生じさせた行動の候補として特別なシンボルを用いて前記学習処理を行い、前記確率P（音響イベント|状況の候補）を得る、モデル処理装置。

【請求項17】

請求項1，2，4の何れかのモデル処理装置であって、

前記音響イベントの時系列の並びは、前記行動ラベル候補によって行動の候補が表された第4音響イベントの時系列の並び、および前記行動ラベル候補によって行動の候補が表されていない第5音響イベントの時系列の並びを含み、

前記モデル化部は、前記第5音響イベントを生じさせた行動の候補として特別なシンボルを用いて前記学習処理を行い、前記確率P（音響イベント|行動の候補）および前記確率P（音響イベント|状況の候補）を得る、モデル処理装置。

【請求項18】

請求項9から17の何れかのモデル処理装置であって、

前記音響信号に対応する音響特徴量を用いて前記第2音響イベントを決定し、前記第2音響イベントラベル列を得る音響イベント判定部を有するモデル処理装置。

【請求項19】

請求項18のモデル処理装置であって、

前記音響信号から音響特徴量を得る特徴量算出部、を有し、

前記音響イベント判定部は、前記特徴量算出部で得られた前記音響特徴量を用いて前記音響イベントを決定し、

前記音響特徴量は、音圧レベル、音響パワー、M F C C 特徴量、L P C 特徴量、立ち上がり特性、調波性、時間周期性の何れかを含む、モデル処理装置。

【請求項20】

少なくとも、各時間区間の音響信号を生じさせた音響イベントの時系列の並びを表す音響イベントラベル列、ならびに、前記音響イベントを生じさせた行動の候補を表す行動ラベル候補および前記音響イベントを生じさせた状況の候補の総数の少なくとも一方を用い、

前記音響イベントの時系列の並びにおける、前記行動の候補に対応する前記音響イベントの組み合わせと前記行動の候補の時系列の並びとの同時事後確率、

前記音響イベントの時系列の並びにおける、前記行動の候補に対応する前記音響イベントの組み合わせと前記行動の候補の時系列の並びと前記行動の候補に対応する前記状況の候補の組み合わせとの同時事後確率、または

10

20

30

40

50

前記音響イベントの時系列の並びにおける、前記状況の候補に対応する前記音響イベントの組み合わせと前記状況の候補の時系列の並びとの同時事後確率の何れかを最大化するための処理を含む処理を行って、行動の候補の何れかおよび状況の候補の何れかの少なくとも一方を選択する分析装置。

【請求項 2 1】

モデル処理装置によって行われるモデル処理方法であって、

少なくとも、各時間区間の音響信号を生じさせた音響イベントの時系列の並びを表す音響イベントラベル列、ならびに、前記音響イベントを生じさせた行動の候補を表す行動ラベル候補および前記音響イベントを生じさせた状況の候補の総数の少なくとも一方を用い、

前記音響イベントの時系列の並びにおける、前記行動の候補に対応する前記音響イベントの組み合わせと前記行動の候補の時系列の並びとの同時事後確率、

前記音響イベントの時系列の並びにおける、前記行動の候補に対応する前記音響イベントの組み合わせと前記行動の候補の時系列の並びと前記行動の候補に対応する前記状況の候補の組み合わせとの同時事後確率、または

前記音響イベントの時系列の並びにおける、前記状況の候補に対応する前記音響イベントの組み合わせと前記状況の候補の時系列の並びとの同時事後確率の何れかを最大化するための処理を含む学習処理によって、行動の候補が音響イベントを生成する確率 P (音響イベント | 行動の候補) および状況の候補が音響イベントを生成する確率 P (音響イベント | 状況の候補) の少なくとも一方を得る、モデル処理方法。

【請求項 2 2】

モデル処理装置によって行われるモデル処理方法であって、

少なくとも、行動の候補が音響イベントを生成する確率 P (音響イベント | 行動の候補) および状況の候補が音響イベントを生成する確率 P (音響イベント | 状況の候補) の少なくとも一方、ならびに、

各時間区間の音響信号を生じさせた第 2 音響イベントの時系列の並びを表す第 2 音響イベントラベル列を用い、

前記第 2 音響イベントの時系列の並びに対する、前記確率 P (音響イベント | 行動の候補) のもとでの、行動の候補の尤もらしさ、

前記第 2 音響イベントの分布と、行動の候補に対応する前記確率 P (音響イベント | 行動の候補) で表される音響イベントの分布との近さ、

前記第 2 音響イベントの時系列の並びに対する、前記確率 P (音響イベント | 状況の候補) のもとでの、状況の候補の尤もらしさ、および

前記第 2 音響イベントの分布と、状況の候補に対応する前記確率 P (音響イベント | 状況の候補) で表される音響イベントの分布との近さ、の少なくとも何れかに基づいて、行動の候補の何れかおよび状況の候補の何れかの少なくとも一方を選択する、モデル処理方法。

【請求項 2 3】

分析装置によって行われる分析方法であって、

少なくとも、各時間区間の音響信号を生じさせた音響イベントの時系列の並びを表す音響イベントラベル列、ならびに、前記音響イベントを生じさせた行動の候補を表す行動ラベル候補および前記音響イベントを生じさせた状況の候補の総数の少なくとも一方を用い、

前記音響イベントの時系列の並びにおける、前記行動の候補に対応する前記音響イベントの組み合わせと前記行動の候補の時系列の並びとの同時事後確率、

前記音響イベントの時系列の並びにおける、前記行動の候補に対応する前記音響イベントの組み合わせと前記行動の候補の時系列の並びと前記行動の候補に対応する前記状況の候補の組み合わせとの同時事後確率、または

前記音響イベントの時系列の並びにおける、前記状況の候補に対応する前記音響イベントの組み合わせと前記状況の候補の時系列の並びとの同時事後確率

10

20

30

40

50

の何れかを最大化するための処理を含む処理を行って、行動の候補の何れかおよび状況の候補の何れかの少なくとも一方を選択する分析方法。

【請求項 24】

請求項 1 から 19 の何れかのモデル処理装置、または請求項 20 の分析装置としてコンピュータを機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、どのような行動および／または状況が音響イベントを生じさせたかを推定または分析するための技術に関する。

10

【背景技術】

【0002】

非特許文献 1 に開示された従来技術では、行動（例えば、人が料理をしている行動等）によって生じた音響信号に対して、短時間（20 msec ~ 100 msec 程度）ごとにその音響信号が何の音（足音、水が流れる音；以後、音響イベントとする）であるかを表すラベルを付与した音響イベントラベル付き音響信号列を入力とし、連続する有限個のフレーム分の音響イベントラベルから音響イベントラベルごとのヒストグラムを作成する。生成された音響イベントラベルごとのヒストグラムに対して GMM (Gaussian Mixture Model)、HMM (Hidden Markov Model)、SVM (Support Vector Machine) 等のモデル化手法を用い、行動を推定するためのモデル（行動モデル）を生成する。

20

【0003】

さらに、上記行動モデルと新たに入力された音響イベントラベル付き音響信号列から算出された音響イベントのヒストグラムをそれぞれ比較し（例えば、ユークリッド距離やコサイン距離などを用いて比較する）、複数の行動モデルのうち、最も判断基準に適合しているものをその音響信号列に対応する行動を表すと判定する。このように、従来技術では音響信号列から行動を推定することができる。

【先行技術文献】

【非特許文献】

【0004】

【非特許文献 1】井本他、「複数の生活音の出現頻度に基づくユーザ行動の識別手法とコミュニケーションへの応用」，画像電子学会第 32 回 VMA 研究会

30

【発明の概要】

【発明が解決しようとする課題】

【0005】

従来技術のモデルには時間情報が考慮されておらず、時間情報によって特徴付けられる行動や音響イベントをモデル化していなかった。これは、時間情報によって特徴付けられた行動の推定精度の劣化につながる。このことは、行動を推定する場合に限られたことではない。行動を分析する場合、または状況を分析もしくは推定する場合にも同様である。

【0006】

本発明は、時間情報を考慮してモデル化を行い、行動および／または状況の推定および／または分析の精度を向上させることを課題とする。

40

【課題を解決するための手段】

【0007】

本発明では、モデルの学習処理に、少なくとも、各時間区間の音響信号を生じさせた音響イベントの時系列の並びを表す音響イベントラベル列、ならびに、音響イベントを生じさせた行動の候補を表す行動ラベル候補および音響イベントを生じさせた状況の候補の総数の少なくとも一方を用いる。この学習処理は、音響イベントの時系列の並びにおける、行動の候補に対応する音響イベントの組み合わせと行動の候補の時系列の並びとの同時事後確率、音響イベントの時系列の並びにおける、行動の候補に対応する音響イベントの組み合わせと行動の候補の時系列の並びと行動の候補に対応する状況の候補の組み合わせと

50

の同時事後確率、または音響イベントの時系列の並びにおける、状況の候補に対応する音響イベントの組み合わせと状況の候補の時系列の並びとの同時事後確率の何れかを最大化するための処理を含む。これにより、行動の候補が音響イベントを生成する確率 P (音響イベント | 行動の候補) および状況の候補が音響イベントを生成する確率 P (音響イベント | 状況の候補) の少なくとも一方を得る。

【発明の効果】

【0008】

本発明では、行動および / または状況の候補の時系列情報を考慮してモデル学習を行うため、行動および / または状況の推定および / または分析の精度を向上できる。

【図面の簡単な説明】

【0009】

【図1】実施例(1)-1-1のモデル処理装置を説明するためのブロック図。

【図2】音響イベントラベルおよび行動ラベル候補付き音響信号列のデータ構成を例示した図。

【図3】実施例(1)-1-2のモデル処理装置を説明するためのブロック図。

【図4】実施例(1)-1-3のモデル処理装置を説明するためのブロック図。

【図5】実施例(1)-2-1のモデル処理装置を説明するためのブロック図。

【図6】実施例(1)-2-2のモデル処理装置を説明するためのブロック図。

【図7】実施例(1)-2-3のモデル処理装置を説明するためのブロック図。

【図8】実施例(1)-3-1のモデル処理装置を説明するためのブロック図。

【図9】実施例(1)-3-2, (1)-4-2のモデル処理装置を説明するためのブロック図。

【図10】実施例(1)-3-3, (1)-4-3のモデル処理装置を説明するためのブロック図。

【図11】実施例(1)-4-1のモデル処理装置を説明するためのブロック図。

【図12】実施例(2)-1-1のモデル処理装置を説明するためのブロック図である。

【図13】音響イベントラベルおよび行動ラベル候補付き音響信号列のデータ構成を例示した図。

【図14】実施例(2)-1-2のモデル処理装置を説明するためのブロック図。

【図15】実施例(2)-1-3のモデル処理装置を説明するためのブロック図。

【図16】実施例(2)-2-1のモデル処理装置を説明するためのブロック図。

【図17】実施例(2)-2-2のモデル処理装置を説明するためのブロック図。

【図18】実施例(2)-2-3のモデル処理装置を説明するためのブロック図。

【図19】実施例(2)-3-1のモデル処理装置を説明するためのブロック図。

【図20】実施例(2)-3-2, (2)-4-2のモデル処理装置を説明するためのブロック図。

【図21】実施例(2)-3-3, (2)-4-3のモデル処理装置を説明するためのブロック図。

【図22】実施例(2)-4-1のモデル処理装置を説明するためのブロック図。

【図23】実施例(3)-1-1のモデル処理装置を説明するためのブロック図。

【図24】音響イベントラベル付き音響信号列のデータ構成を例示した図。

【図25】実施例(3)-1-2のモデル処理装置を説明するためのブロック図。

【図26】実施例(3)-1-3のモデル処理装置を説明するためのブロック図。

【図27】実施例(3)-2-1のモデル処理装置を説明するためのブロック図。

【図28】実施例(3)-2-2のモデル処理装置を説明するためのブロック図。

【図29】実施例(3)-2-3のモデル処理装置を説明するためのブロック図。

【図30】実施例(3)-3-1のモデル処理装置を説明するためのブロック図。

【図31】実施例(3)-3-2のモデル処理装置を説明するためのブロック図。

【図32】実施例(3)-3-3のモデル処理装置を説明するためのブロック図。

【0010】

10

20

30

40

50

以下、図面を参照して本発明の実施形態を説明する。なお、同一構成要素ないし同一処理には同一符号を割り当てて重複説明を省略する。

<用語の定義>

各実施例で用いる用語を定義する。

「行動」とは、人間、動物、装置などの主体が行う何らかの行動を意味する。「行動」の具体例は「料理」「掃除」などである。また、音響信号が収録された場所や時間や場に人が何人いるか等の音響信号を発生させる状態を「行動」としてもよい。

「音響イベント」とは、行動に伴って発生する音の事象を意味する。「音響イベント」の具体例は、「包丁の音」「水が流れる音」「水音」「着火音」「火の音」「足音」「掃除機の排気音」などである。例えば、「包丁の音」「水が流れる音」等は、「料理」という行動に伴って発生する音響イベントである。10

「状況」とは、音響イベントの組み合わせ、および状況そのものの遷移確率で特徴付けられた状況の種別（番号）である。言い換えると「状況」は、音響イベントおよび自身の遷移確率によって規定される潜在的な場の状況を表す。状況の生成確率は、その状況が起こる時間区間での行動または、その状況が起こる時間区間での音響信号そのものに規定され、状況は、その状況が起こる時間区間での音響イベントの生成確率を規定する。状況は、行動と、行動によって規定される生成確率、または、音響信号と、音響信号によって規定される生成確率とによって表現可能である。音響イベントは、行動と、行動によって規定される生成確率、または、状況と、状況によって規定される生成確率とによって表現可能である。なお、行動や状況の推定は長時間の音響イベントの組み合わせによって実現される。20

「XがYを生成する確率」とは、事象Xが起こるという条件のもとでの事象Yが起こる確率をいう。「XがYを生成する確率」は、「XのもとでのYの条件付き確率」や「XにおけるYの条件付き確率」とも表現できる。「Xにおける、YとZとの同時事後確率」とは、事象Xが起きたという条件のもとでの、事象Yと事象Zとの同時確率をいう。

【0011】

<実施例(1)-1-1：図1>

[行動の時間変化を考慮に入れた、音響イベントラベルおよび行動ラベル候補付き音響信号列から行動・音響イベント生成モデルおよび行動遷移モデルを算出し、同時に、行動を分析]30

本実施例では、行動の時間変化を考慮に入れ、音響イベントラベルおよび行動ラベル候補付き音響信号から、行動・音響イベント生成モデルを算出する。同時に、行動の時間遷移を表す行動遷移モデルも算出してもよく、行動の分析も行ってもよい。

【0012】

すなわち本実施例では、各時間区間の音響信号を生じさせた音響イベントの時系列の並びを表す音響イベントラベル列、ならびに、音響イベントを生じさせた行動の候補を表す行動ラベル候補を用い、「音響イベントの時系列の並びにおける、行動の候補に対応する音響イベントの組み合わせと行動の候補の時系列の並びとの同時事後確率」を最大化するための処理を行う学習処理を行う。例えば、この学習処理は、行動の時間遷移に基づく行動の出現確率と、行動に対する音響イベントの出現確率に基づいて、音響イベントの時系列の並びにおける、行動の候補に対する音響イベントの組み合わせと行動の候補の時系列の並びとの同時事後確率を最大化するための処理である。例えば、音響イベントの時系列の並びからなる音響イベントの組み合わせに対する尤度関数P（音響イベントの組み合わせ | 行動の候補の時系列の並び）を最大にする処理が行われる。それにより、行動の候補が音響イベントを生成する確率P（音響イベント | 行動の候補）を「行動・音響イベントの生成モデル」として得る。また、この学習処理によって、行動の候補の時系列の遷移確率を「行動遷移モデル」として得てもよい。さらに、この学習処理の過程で何れかの行動の候補を選択し、各音響イベントがどの行動によって生成されたのかを分析してもよい。この場合には選択した行動の候補を表す行動ラベルを出力してもよい。40

【0013】

10

20

30

40

50

図1に例示するように、本実施例のモデル処理装置100は、音響信号列合成部101、モデル化部102、および記憶部103を有する。モデル処理装置100は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0014】

まず音響信号列合成部101に、音響イベントラベルおよび行動ラベル候補付き音響信号列 $11 - 1, \dots, 11 - S$ （ただし、Sは1以上の整数）が入力される。図2に例示するように、音響イベントラベルおよび行動ラベル候補付き音響信号列は、短時間（数 10 msec ～数 sec ）の時間区間（以下単に「時間区間」という）ごとに区分された時系列の音響信号列、当該時間区間ごとに区分された音響信号列の各要素に対応する要素番号、当該時間区間ごとに決定されて付与された音響イベントラベル、および当該時間区間ごとに付与された行動ラベル候補を含む。当該時間区間ごとに複数個の行動ラベル候補が付与されてもよいし（例えば、図2の要素番号1の時間区間）、単一の行動ラベル候補が付与されてもよい（例えば、図2の要素番号iの時間区間）。音響信号列は音を表すデジタル信号列（例えば、5～20秒程度のデジタル信号列）である。音響信号列は、例えば单一又は複数の行動がなされる環境での録音によって得られ、单一又は複数の行動によって生じた音を表す。音響信号列の各時間区間の区分を音響信号とよぶ。音響信号列は、各時間区間の音響信号の時系列の並びと解釈できる。音響イベントラベルは、それぞれ、上記の時間区間の音響信号を生じさせた音響イベントを表す。各時間区間の音響イベントラベルの時系列の並びを音響イベントラベル列と呼ぶ。音響イベントラベル列は、各時間区間の音響信号を生じさせた音響イベントの時系列の並びを表す。行動ラベル候補は、音響イベントを生じさせた行動の候補を表すラベルである。

【0015】

複数個の音響イベントラベルおよび行動ラベル候補付き音響信号列 $11 - 1, \dots, 11 - S$ （以下、単に「ラベル付き音響信号列 $11 - 1, \dots, 11 - S$ 」という）が音響信号列合成部101に入力された場合、音響信号列合成部101は、それらを時系列方向につなぎ合わせ、それによって1つの音響イベントラベルおよび行動ラベル候補付き音響信号列 11 （以下、単に「ラベル付き音響信号列 11 」という）を得て出力する（合成処理）。音響信号列合成部101に1つの音響信号列 $11 - 1$ のみが入力された場合、音響信号列合成部101はそれをラベル付き音響信号列 11 として出力する。音響信号列合成部101から出力された音響イベントラベルおよび行動ラベル候補付き音響信号列は、モデル化部102に入力される。なお、音響信号列合成部101を経由することなく、1つラベル付き音響信号列 11 がそのままモデル化部102に入力されてもよい。

【0016】

モデル化部102は、以下の手順に従って、入力されたラベル付き音響信号列 11 から、行動-音響イベント生成モデル14を生成し、それを記憶部103に格納する。モデル化部102は、行動遷移モデルを生成してもよいし、行動ラベル18を生成してもよい。生成された行動遷移モデル17や行動ラベル18も記憶部103に格納される。

【0017】

[行動から音響イベントが生成される過程の論理的説明]

モデル化部102は、ラベル付き音響信号列 11 から、行動の候補が音響イベントを生成する確率P（音響イベント|行動の候補）としたときの行動-音響イベント生成モデル14を得る。このとき、行動が音響イベントの生成確率を規定するとともに、ある時刻の行動が次時刻に生じる行動を規定すると考え、この関係を生成モデルとして記述する。この際、行動-音響イベント生成モデル14のみを得てもよいし、行動の候補の時系列の遷移確率である行動遷移モデル17を同時に得てもよい。

【0018】

ラベル付き音響信号列 11 が有する各音響信号列が取りうる行動ラベル候補を音響信号列分並べて列とした、集合のベクトル、行動ラベル候補が表す行動の候補における音響イベントの生成確率、および行動ラベル候補が表す行動の候補の遷移確率が与えられ

10

20

30

40

50

た場合における、音響イベントの列 e の生成確率 $P(e | \cdot, \cdot, \cdot)$ は以下の通りである。

【数1】

$$P(e | \Pi, \Phi, \Omega) = \prod_{s=1}^S P(e'_s | \Pi, \Phi, a_s) \quad (1)$$

ただし、 A は行動の候補の総数（行動の候補の種類の個数）、 E は音響イベントの種類の個数、 e はラベル付き音響信号列 11 に与えられた音響イベントの列（ベクトル）、 \cdot は各時間区間の行動の候補 a （ただし、 $a \in \{1, \dots, A\}$ ）が次の時間区間で行動の候補 a' （ただし、 $a' \in \{1, \dots, A\}$ ）に遷移する確率を (a, a') 要素とする $A \times A$ 行列、 \cdot は行動の候補 a （ただし、 $a \in \{1, \dots, A\}$ ）が音響イベント（ただし、 $\cdot \in \{1, \dots, E\}$ ）を生成する確率 $P(\cdot | a)$ を (a, \cdot) 要素とする $A \times E$ 行列、 \cdot は各音響信号列がとり得る行動ラベル候補を音響信号列分並べて列とした、集合のベクトル、 e'_s は各ラベル付き音響信号列 11 - s に与えられた音響イベントの列（ N_s 次元ベクトル： N_s はラベル付き音響信号列 11 - s に対応する音響イベントの個数）、 a_s はラベル付き音響信号列 11 - s に付与された行動ラベル候補が表す行動の候補の集合、 S はラベル付き音響信号列 11 を構成するラベル付き音響信号列 11 - s の個数を表す。なお、 \cdot と a_s の関係は以下のように表される。

$$= [a_1, a_2, \dots, a_s, \dots, a_S]$$

【0019】

行動の候補の集合 a_s 、行動の候補 a の遷移確率 \cdot_a の事前分布のパラメータ（Dirichlet 分布に従うものとする）、行動の候補 a における音響イベント \cdot の生成確率 \cdot_a の事前分布のパラメータ（Dirichlet 分布に従うものとする）が与えられたときの、音響イベントの列 e'_s の生成確率 $P(e'_s, \cdot, \cdot | \cdot, \cdot, a_s)$ は、以下の通りである。

【数2】

$$P(e'_s | \Pi, \Phi | \gamma, \beta, a_s)$$

$$= P(\pi_a | \gamma) P(\phi_a | \beta) \prod_{i=1}^{N_s} P(e_i | x_i, \phi_a) P(x_i | \pi_{x_i}^{x_{i-1}}, a_s) \quad (2)$$

ただし、 e_i はラベル付き音響信号列 11 - s の要素番号 i に対応する音響イベント、 N_s はラベル付き音響信号列 11 - s に対応する要素数（要素番号 i の最大値）、 x_i はラベル付き音響信号列 11 - s の要素番号 i に対応する行動ラベル候補が表す行動の候補、 $x_{i-1} x_i$ はラベル付き音響信号列 11 - s の要素番号 $i - 1$ に対応する行動ラベル候補が表す行動の候補 x_{i-1} から要素番号 i に対応する行動ラベル候補が表す行動の候補 x_i への遷移確率を表す。ただし、 $x_{i-1} x_i$ の上付き「 x_{i-1} 」は x_{i-1} を表し、下付き「 x_i 」は x_i を表す。また、 \cdot_a は、ラベル付き音響信号列 11 に付与されたすべての行動ラベル候補を対象とした遷移確率 $\cdot_{x_{i-1} x_i}$ を (x_{i-1}, x_i) 要素とする $A \times A$ 行列、 \cdot_a は行動の候補 a が音響イベント（ただし、 $\cdot \in \{1, \dots, E\}$ ）を生成する確率 $P(\cdot | a)$ を \cdot 番目の要素とする E 次元ベクトルを表す。W - 1 次（Wは2以上の整数）の Dirichlet 分布の確率密度関数は以下の通りである。

【数3】

$$\text{Dir}(\mu | \tau) = \frac{\Gamma(\sum_{i=1}^W \tau_i)}{\prod_{i=1}^W \Gamma(\tau_i)} \prod_{i=1}^W \mu^{\tau_i - 1}$$

10

20

30

40

50

ただし、 θ_i ($i = 1, \dots, W$) からなるパラメータ、 μ は確率変数、 γ はガンマ関数を表す。

【0020】

[生成モデルの算出過程の説明]

モデル化部 102 は、入力されたラベル付き音響信号列 11 から行動の候補が音響イベントを生成する確率 P (音響イベント | 行動の候補) および行動の遷移確率を算出する。また、モデル化部 102 は、少なくとも行動 - 音響イベント生成モデル 14 を出力し、記憶部 103 に格納する。また、モデル化部 102 は、算出された行動の候補の遷移確率を行動遷移モデル 17 として記憶部 103 に格納してもよい。行動の候補の遷移確率は、第 1 時間区間での行動の候補が、第 1 時間区間よりも後の第 2 時間区間での行動の候補を生成する確率 P (第 2 時間区間での行動の候補 | 第 1 時間区間での行動の候補) である。例えば、第 2 時間区間は第 1 時間区間の直後の時間区間である。しかしながら、第 1 時間区間よりも k 個後 (k は 2 以上の整数) の時間区間が第 2 時間区間であってもよい。また、第 1 時間区間から第 k 時間区間の同時確率を考えてもよい。さらにモデル化部 102 は、入力されたラベル付き音響信号列 11 が有する行動ラベル候補から尤もらしいラベルを選択し、選択した行動ラベル候補からなる行動ラベル列 18 を外部に出力したり、記憶部 103 に格納したりしてもよい。尤もらしいラベルは、各時間区間にに対して 1 個でもよいし、複数個でもよい。また、各時間区間ににおいて尤もらしさが閾値を超えたすべての行動ラベル候補を選択し、選択した時間区間ごとに単一または複数の行動ラベル候補からなる行動ラベル 18 の列を外部に出力したり、記憶部 103 に格納したりしてもよい。

【0021】

行動 - 音響イベント生成モデル 14 および行動遷移モデル 17 の算出は、入力されたラベル付き音響信号列 11 の音響イベントの時系列の並びにおける、行動の候補に対応する音響イベントの組み合わせと行動の候補の時系列の並びとの同時事後確率を最大化する処理を用いて行われる。例えば、マルコフ連鎖モンテカルロ法 (MCMC 法 : Markov Chain Monte Carlo methods) や変分ベイズ法 (VB 法 : Variational Bayes methods) などの手法を用いることができる。MCMC 法には、M-H アルゴリズムやギブスサンプリングなどの手法がある。ここでは一例としてギブスサンプリングによる算出手法を説明する。

【0022】

[生成モデル算出の流れの例示]

(i) モデル化部 102 は、ハイパーパラメータとしてパラメータ θ および γ を事前に決定する。

(ii) モデル化部 102 は、全てのラベル付き音響信号列 11 - s (ただし $s = 1, \dots, S$) の全ての要素番号 i に対応する要素に行動の候補 x_i をランダムに割り当てる (割り当てはラベル付き音響信号列 11 - s に与えられた行動ラベル候補が表す行動の候補からランダムに行う)。あるいは、要素番号 i に対応する要素への行動の候補 x_i の割り当ては、事前にモデル化部 102 に設定しておいた割り当てであってもよい。

【0023】

(iii) モデル化部 102 は、全てのラベル付き音響信号列 11 - s (ただし $s = 1, \dots, S$) の各要素番号 i (要素番号 $i - 1$ または $i + 1$ に対応する要素が存在しない要素番号 i を除く) について、以下の(iii-1)および(iii-2)を規定の回数 (正值であり、例えば 1 ~ 3000 回程度)、もしくは、所望の結果が得られるまで (例えば、割り当ての前後において、行動の割り当て先の変化が一定の閾値 (例えば 0.3 %) 以下になるまでなど) 繰り返す。

(iii-1) モデル化部 102 は、以下の式により、更新対象のラベル付き音響信号列の要素番号 i を除いた要素番号に対応する音響信号列に対して、要素番号 i の要素に行動の候補 a に割り当てられる確率分布 $P(x_i = a | e_i = \dots, x_{-i}, e_{-i}, \dots)$ を更新する。

【数4】

$$P(x_i = a | \varepsilon_i = \varepsilon, x_{-i}, \varepsilon_{-i}, \Omega, \beta, \gamma)$$

$$\propto \frac{C_{ea}^{EA} + \beta}{(\sum_{\varepsilon'=1}^E C_{\varepsilon'a}^{EA}) + E\beta}$$

$$\times \frac{(R_{x_i}^{x_{i-1}} + \gamma)}{\left\{ (\sum_{x'_i \in \Omega} R_{x'_i}^{x_{i-1}}) + A\gamma \right\}}$$

$$\times \frac{(R_{x_{i+1}}^{x_i} + I(x_{i-1} = x_i) \times I(x_i = x_{i+1}) + \gamma)}{\left\{ (\sum_{x'_{i+1} \in \Omega} R_{x'_{i+1}}^{x_i}) + I(x_{i-1} = x_i) + A\gamma \right\}}$$

$$\propto \frac{C_{ea}^{EA} + \beta}{(\sum_{\varepsilon'=1}^E C_{\varepsilon'a}^{EA}) + E\beta}$$

$$\times \frac{(R_{x_i}^{x_{i-1}} + \gamma)(R_{x_{i+1}}^{x_i} + I(x_{i-1} = x_i) \times I(x_i = x_{i+1}) + \gamma)}{\left\{ (\sum_{x'_{i+1} \in \Omega} R_{x'_{i+1}}^{x_i}) + I(x_{i-1} = x_i) + A\gamma \right\}} \quad (3)$$

ただし、 $C_{a}^{E\text{EA}}$ は、更新対象のラベル付き音響信号列（初期値はラベル付き音響信号列 11）の音響イベント に行動の候補 a が割り当てられた回数を表す。

【0024】

$R_{x_i}^{x_{i-1} x_{i+1}}$ は、更新対象のラベル付き音響信号列に対応するすべて行動の候補の並びの中で、行動の候補 $x_i = a$ と同一の候補から行動の候補 x_{i+1} と同一の候補に遷移した回数を表す。ただし、 $x_i = a$ であり、 x_i 以外の行動の候補「 $\dots, x_{i-1}, x_{i+1}, \dots$ 」は、(ii)でランダムに割り当てられた行動の候補 [(iii-2) が実行されていない場合] または直前の(iii-2)で得られた行動の候補 [(iii-2) が実行されている場合] である。 x_{i-1} から x_i への遷移「 $x_{i-1} x_i$ 」および x_i から x_{i+1} への遷移「 $x_i x_{i+1}$ 」は $R_{x_i}^{x_{i-1} x_{i+1}}$ を得るためのカウント対象とせず、遷移「 $x_{i-1} x_i$ 」が遷移「 $a x_{i+1}$ 」と同じになることは $I(x_{i-1} = x_i) \times I(x_i = x_{i+1})$ によってカウントされる。ただし、 $I(x_{i-1} = x_i)$ および $I(x_i = x_{i+1})$ は、 $x_{i-1} = x_i$ のときに $I(x_{i-1} = x_i) = 1$ 、 $x_{i-1} \neq x_i$ のときに $I(x_{i-1} = x_i) = 0$ となり、 $x_i = x_{i+1}$ のときに $I(x_i = x_{i+1}) = 1$ 、 $x_i \neq x_{i+1}$ のときに $I(x_i = x_{i+1}) = 0$ となる関数である。遷移「 $x_{i-1} x_i$ 」が遷移「 $a x_{i+1}$ 」と同じになることが $I(x_{i-1} = x_i) \times I(x_i = x_{i+1})$ によってカウントできるのは、遷移「 $x_{i-1} x_i$ 」が遷移「 $a x_{i+1}$ 」と同じになるのは、 $a = x_i = x_{i+1}$ かつ $a = x_{i-1} = x_i$ の場合に限られるからである。

【0025】

例えば、 $= \{1, 2, \dots, 9\}$ であり、前回の繰り返し処理時に得られた各要素番号 i に対応する行動の候補の並びが $[x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8] = [1, 2, 1, 1, 2, 3, 1, 6]$ であったとする。この例で $i = 2$ について $R_{x_i}^{x_{i-1} x_{i+1}}$ を求める場合には、 $x_2 = a = \{1, 2, \dots, 9\}$ とした $[1, a, 1, 1, 2, 1, 1, 6]$ の中で「 $a = 1$ 」と遷移する回数をカウントする。ただし、遷移「 $x_1 x_2$ 」=「 $1 a$ 」および「 $x_2 x_3$ 」=「 $a 1$ 」は $R_{x_i}^{x_{i-1} x_{i+1}}$ を得るためのカウント対象としない。

10

20

30

40

50

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
1	a	1	1	2	1	2	6
1							
2							
...							
9							

$x_2 = a = 1$ の場合、「 $a = 1$ 」は「 $1 = 1$ 」であり、 $[1, 1, 1, 1, 2, 1, 2, 6]$ の中で「 $1 = 1$ 」と遷移するのは、「 $x_1 \times x_2$ 」=「 $1 = 1$ 」、「 $x_3 \times x_4$ 」=「 $1 = 1$ 」の3回である。しかしながら、「 $x_1 \times x_2$ 」=「 $1 = 1$ 」および「 $x_2 \times x_3$ 」=「 $1 = 1$ 」は $R^{\times i_{x_i+1}}$ を得るためのカウント対象とせず、 $a = 1$ および $i = 2$ について $R^{\times i_{x_i+1}} = 1$ とする。なお「 $x_1 \times x_2$ 」=「 $1 = 1$ 」が「 $x_1 \times x_2$ 」=「 $1 = 1$ 」と一致することは、 $I(x_1 = x_2 = 1) \times I(x_2 = x_3 = 1) = 1$ によってカウントされる。 $a = 2$ の場合、「 $a = 1$ 」は「 $2 = 1$ 」であり、「 $1, 2, 1, 1, 2, 1, 2, 6$ 」の中で「 $2 = 1$ 」と遷移るのは、「 $x_2 \times x_3$ 」=「 $2 = 1$ 」、「 $x_5 \times x_6$ 」=「 $2 = 1$ 」の2回である。しかしながら、「 $x_2 \times x_3$ 」=「 $2 = 1$ 」は $R^{\times i_{x_i+1}}$ を得るためのカウント対象とせず、 $a = 2$ および $i = 2$ について $R^{\times i_{x_i+1}} = 1$ とする。なお「 $x_1 \times x_2$ 」=「 $1 = 1$ 」については、 $x_1 \times x_2 = 2$ かつ $2 = x_2 \times x_3$ であるため、 $I(x_1 = x_2) \times I(x_2 = x_3) = 0$ となり、カウントされない。

【0026】

10

$x_{i+1} R^{\times i_{x_{i+1}}}$ は、更新対象のラベル付き音響信号列に対応するすべて行動の候補の並びの中で、行動の候補 $x_i = a$ と同一の候補から遷移した行動の候補の総数を表す。この場合も、 x_{i-1} を基点とする遷移および $x_i = a$ を基点とする遷移は、 $x_{i+1} R^{\times i_{x_{i+1}}}$ を得るためのカウント対象とせず、 x_{i-1} を基点とする遷移が $x_i = a$ を基点とする遷移となることは、 $I(x_{i-1} = x_i)$ によってカウントされる。 x_{i-1} を基点とする遷移が $x_i = a$ を基点とする遷移となることが $I(x_{i-1} = x_i)$ によってカウントできるのは、 x_{i-1} を基点とする遷移が $x_i = a$ を基点とする遷移となるのは $x_{i-1} = a$ の場合、すなわち $x_{i-1} = x_i = a$ の場合だからである。

20

例えば、上述の行動の候補の並びの例 $[x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8] = [1, 2, 1, 1, 2, 3, 1, 6]$ において、 $i = 2$ について $x_{i+1} R^{\times i_{x_{i+1}}}$ を求める場合には、 $x_2 = a \{1, 2, \dots, 9\}$ とした $[1, a, 1, 1, 2, 1, 1, 6]$ の中で「 a 」を基点とする遷移回数をカウントする。ただし、遷移「 $x_1 \times x_2$ 」=「 $1 = a$ 」および「 $x_2 \times x_3$ 」=「 $a = 1$ 」は $x_{i+1} R^{\times i_{x_{i+1}}}$ を得るためのカウント対象としない。 $a = 1$ の場合、「 $1, 1, 1, 1, 2, 1, 2, 6$ 」の中で $x_2 = 1$ を基点として遷移するのは、「 $x_1 \times x_2$ 」=「 $1 = 1$ 」、「 $x_2 \times x_3$ 」=「 $1 = 1$ 」、「 $x_3 \times x_4$ 」=「 $1 = 1$ 」、「 $x_6 \times x_7$ 」=「 $1 = 1$ 」の4回である。しかしながら、「 $x_1 \times x_2$ 」=「 $1 = 1$ 」および「 $x_2 \times x_3$ 」=「 $1 = 1$ 」は $x_{i+1} R^{\times i_{x_{i+1}}}$ を得るためのカウント対象とせず、 $a = 1$ および $i = 2$ について $x_{i+1} R^{\times i_{x_{i+1}}} = 2$ とする。なお、 x_1 を基点とする遷移が $x_2 = 1$ を基点とする遷移となることは、 $I(x_1 = x_2 = 1) = 1$ によってカウントされる。

30

$a = 2$ の場合、「 $1, 2, 1, 1, 2, 1, 2, 6$ 」の中で $x_2 = 2$ を基点として遷移するのは、「 $x_2 \times x_3$ 」=「 $2 = 1$ 」、「 $x_5 \times x_6$ 」=「 $2 = 1$ 」、「 $x_7 \times x_8$ 」=「 $2 = 6$ 」の3回である。しかしながら、「 $x_2 \times x_3$ 」=「 $2 = 1$ 」は $x_{i+1} R^{\times i_{x_{i+1}}}$ を得るためのカウント対象とせず、 $a = 2$ および $i = 2$ について $x_{i+1} R^{\times i_{x_{i+1}}} = 2$ とする。なお「 $x_1 \times x_2$ 」=「 $1 = 1$ 」については、「 $x_1 \times x_2 = 2$ 」であるため、 $I(x_1 = x_2) = 0$ となり、カウントされない。

【0027】

なお、表記制約上の都合から「 $C_a^{E_A}$ 」「 $R^{\times i_{x_{i+1}}}$ 」と表記するが、本来は

50

式(3)に示すように「 $C_{a \rightarrow E^A}$ 」の「EA」は「a」の上に表記され、「 $R^{x_i \times x_{i+1}}$ 」の「 x_i 」は x_i を表し、「 x_{i+1} 」は x_{i+1} を表し、 x_i は x_{i+1} の上に表記される。また、 x_{-i} は要素番号i以外の要素番号に対応する行動の候補からなる列、 e_{-i} は要素番号i以外の要素番号に対応する音響イベントからなる列を表す。

【0028】

(iii-2)モデル化部102は、上記の更新式(3)で得られた確率分布 $P(x_i = a | e_i = \cdot, x_{-i}, e_{-i}, \cdot, \cdot, \cdot)$ に従って、各要素番号iの要素に割り当てる行動の候補 x_i をサンプリング(選択)する。モデル化部102は、このようにサンプリングした行動の候補 x_i を要素番号iの要素に割り当て、要素番号iに対応する行動の候補を更新する。これにより更新対象のラベル付き音響信号列を更新する。

10

【0029】

(iv)モデル化部102は、上記の(iii-1)(iii-2)の繰り返しによって最終的に得られた $C_{a \rightarrow E^A}$ および $R^{x_i \times x_{i+1}}$ を用い、以下を計算する。

【数5】

$$\phi_{ea} = \frac{C_{ea}^{EA} + \beta}{(\sum_{e'=1}^E C_{e'a}^{EA}) + E\beta} \quad (4)$$

$$\pi_{x_{i+1}}^{x_i} = \frac{(R_{x_i}^{x_{i-1}} + \gamma)(R_{x_{i+1}}^{x_i} + \gamma)}{\{(\sum_{x'_i \in \Omega} R_{x'_i}^{x_{i-1}}) + A\gamma\} \{(\sum_{x'_{i+1} \in \Omega} R_{x'_{i+1}}^{x_i}) + A\gamma\}} \quad (5)$$

20

【0030】

これによってモデル化部102は、行動の候補が音響イベントを生成する確率 ϕ_a (ただし、 $a \in \{1, \dots, A\}$, $\epsilon \in \{1, \dots, E\}$)の集合を得、それを行動-音響イベント生成モデル14とする。例えばモデル化部102は、確率 ϕ_a を(a, ϵ)要素とする $A \times E$ 行列を行動-音響イベント生成モデル14とする。また、モデル化部102は、これによって行動の候補の遷移確率 $\pi_{x_i, x_{i+1}}^{x_{i-1}}$ (ただし、 $x_i, x_{i+1} \in \{1, \dots, A\}$)の集合を得、それを行動遷移モデル17とする。例えばモデル化部102は、 $x_i, x_{i+1} \in \{1, \dots, A\}$ に対応する遷移確率 $\pi_{x_i, x_{i+1}}^{x_{i-1}}$ を(x_i, x_{i+1})要素とする $A \times A$ 行列を行動遷移モデル17とする。

30

【0031】

或いは、モデル化部102は、(iii-1)の繰り返し処理時に、式(3)の算出過程で確率 ϕ_a を1個以上サンプリング(計算)し、式(4)に代えて、サンプリングされた確率 ϕ_a の平均値を用い、行動-音響イベント生成モデル14を得てもよい。同様に、モデル化部102は、(iii-1)の繰り返し処理時に、式(3)の算出過程で遷移確率 $\pi_{x_i, x_{i+1}}^{x_{i-1}}$ を1個以上サンプリングし、式(5)に代えて、サンプリングされた遷移確率 $\pi_{x_i, x_{i+1}}^{x_{i-1}}$ の平均値を用い、行動遷移モデル17を得てもよい。

【0032】

また、上記更新過程でサンプリング(選択)し、各要素番号iの要素に割り当てる行動の候補 x_i を分析することで、各ラベル付き音響信号列11-sの音響信号列がどの行動によって生成されたかを知ることができる。モデル化部102は、このように各要素番号iの要素に割り当てる行動の候補 x_i を表すラベル、またはそれらの一部を行動ラベル18として出力してもよい。

40

【0033】

本実施例では、行動が音響イベントを生成する確率のみではなく、行動の時間遷移確率も考慮して学習処理を行うことで、遷移確率によって特徴付けられる行動も精度よくモデル化できる。また、遷移確率によって特徴付けられる行動を精度よく分析できる。

【0034】

<実施例(1)-1-2:図3>

50

[行動ラベル候補付き音響信号列から行動 - 音響イベント生成モデルおよび行動遷移モデルを算出し、行動を分析]

本実施例では、行動ラベル候補付き音響信号 $15 - 1, \dots, 15 - S$ を入力とし、行動の時間変化を考慮に入れ、行動 - 音響イベント生成モデル 14 を算出する。すなわち本実施例では、音響信号から音響特徴量を得、得られた音響特徴量を用いて音響イベントを決定し、それによって音響イベントラベル列を得てから学習処理を行い、行動 - 音響イベント生成モデル 14 を生成する。なお、この際、行動の時間遷移を表す行動遷移モデル 17 も算出してもよく、行動の分析を行ってもよい。

【0035】

図 3 に例示するように、本実施例のモデル処理装置 110 は、特徴量算出部 111、音響イベント判定部 112、音響イベントモデルデータベース (DB) 113、音響信号列合成部 101、モデル化部 102、および記憶部 103 を有する。モデル処理装置 110 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0036】

まず特徴量算出部 111 に行動ラベル候補付き音響信号列 $15 - 1, \dots, 15 - S$ が入力される。各行動ラベル候補付き音響信号列 $15 - s$ (ただし、 $s \in \{1, \dots, S\}$) は、前述の時間区間ごとに区分された時系列の音響信号列、当該時間区間ごとに区分された音響信号列の各要素に対応する要素番号、および当該時間区間ごとに付与された行動ラベル候補を含む。要素番号および行動ラベル候補は、音響信号列の要素ごとに付与される。1つの時間区間に複数種類の行動ラベル候補が対応してもよいし、1つの時間区間に複数種類の行動ラベル候補が対応してもよい。

【0037】

特徴量算出部 111 は、各行動ラベル候補付き音響信号列 $15 - s$ が含む音響信号列から、各行動ラベル候補付き音響信号列 $15 - s$ に対応する音響特徴量列 (ベクトル) を算出して出力する。例えば特徴量算出部 111 は、入力された行動ラベル候補付き音響信号 $15 - s$ が含む音響信号列に対し、前述の各時間区間であるフレームごとに、音圧レベル、音響パワー、MFCC (Mel-Frequency Cepstrum Coefficient) 特徴量、LPC (Linear Predictive Coding) 特徴量などを算出し、これらを音響特徴量列として出力する。さらに立ち上がり特性、調波性、時間周期性など (例えば、非特許文献 1 参照) の音響特徴量が音響特徴量列に加えられてもよい。

【0038】

立ち上がり特性とは、数十から数百ミリ秒ごとにおける、音響信号の大きさを表す指標の増加の度合いを表す指標である。ここで、音響信号の大きさを表す指標とは、例えば、音響信号の振幅の絶対値、音響信号の振幅の絶対値の対数値、音響信号のパワー又は音響信号のパワーの対数値である。例えば、以下の式 (10) で得られる値が 0 以上であればその値が立ち上がり特性とされ、式 (10) で得られる値が 0 未満であれば 0 が立ち上がり特性とされる。

【数6】

$$\frac{\sum_{k=1}^K \Delta \bar{p}_k \times (\bar{p}_k)^2}{\sum_{k=1}^K \Delta \bar{p}_k} \quad (10)$$

ただし、 k はフレームを K 個の微小な時間区間 (例えば 1 msec 程度) に区分した場合の各時間区間に對応し、 \bar{p}_k は k 番目の時間区間でのサンプルの大きさを表す指標の代表値又は平均値を表す。なお、「サンプルの大きさを表す指標」の例は、サンプルの振幅、サンプルの振幅の絶対値、サンプルの振幅の対数値、サンプルのエネルギー、サンプル

10

20

30

40

50

のパワー、又はサンプルのパワーの対数値などである。「サンプル」は音響信号列の各音響信号を表す。また、 p_{-k} は p_k の変化率を表す。例えば、 $p_{-k} = p_{-k} - p_{-k-1}$ である。 $p_{-k} = p_{-k+1} - p_{-k}$ としてもよい。また、最小二乗法等の近似手法を用いて k 番目の時間区間ににおける p_{-k} を近似した直線を求め、その時間区間ににおけるその直線の傾きを p_{-k} としてもよい。また、 k 番目の時間区間を含む複数の時間区間ににおける $p_{-k}, \dots, p_{-k-1}, p_{-k}, p_{-k+1}, \dots, p_{-k-L}$ の近時曲線を求め、その k 番目の時間区間に對応する点での傾き(微分値)を p_{-k} としてもよい。また \cdot を任意の文字として、 \cdot の右肩の「-」は、 \cdot の上付きバーを意味する。また式(10)の分子における $(p_n)^2$ を $(p_n)^m$ とし、 m を任意の値としても良い。

10

【0039】

以下に調波性を例示する。

【数7】

$$(調波性) = \begin{cases} f(n) = \delta\{x(n) \times x(n-1)\} & (1 \leq n \leq N) \\ \max\{R_{ff}(\tau)\} & (1 \leq \tau \leq N) \end{cases} \quad (11)$$

ただし、

$$\delta\{x(n) \times x(n-1)\} = \begin{cases} 1 & (x(n) \times x(n-1) \leq 0) \\ 0 & (x(n) \times x(n-1) > 0) \end{cases}$$

20

また、 N はフレームに含まれるサンプル数を表す 1 以上の整数、 n はフレーム内の各サンプル点を表す 1 以上の N 以下の整数、 $x(n)$ はサンプル点 n でのサンプルの大きさを表す指標である。 $R_{ff}(\cdot)$ は $f(n)$ のラグ \cdot での自己相関係数、 $\max\{\cdot\}$ は「 \cdot 」の最大値を表す。ラグ \cdot は 1 以上 N 以下の整数である。 $R_{ff}(\cdot)$ は、例えば以下のように定義される。

【数8】

$$R_{ff}(\tau) = \frac{1}{N} \sum_{n=1}^N \frac{f(n) \times f(n+\tau)}{\{f(n)\}^2}$$

30

【0040】

以下に時間周期性を例示する。

【数9】

$$(時間周期性) = \max_L \frac{\sum_{n=1}^N \prod_{m=1}^M p(n+mL)}{\sum_{n=1}^N (\bar{p})^M} \quad (12)$$

ただし、 L は一周期とみなすサンプル数、 M は時間周期性の度合を計算するための周期数を表す 1 以上の整数、 $p(\cdot)$ はサンプルの大きさを表す指標を時間平滑化した値、 \bar{p} はフレーム内でのサンプルの大きさを表す指標の平均値を表す。

40

【0041】

音響イベントモデル D B 1 1 3 には、事前に算出された音響イベントモデルが複数保存されている。各音響イベントモデルは、音響イベントラベルが付された学習用の音響信号列から音響特徴量列を算出し、各音響イベントに対応する音響特徴量列を GMM, HMM, SVM 等の周知のモデル化手法を用いてモデル化することで得られる(例えば参考文献: 奥村学、高村大也、「言語処理のための機械学習入門」コロナ社)。

【0042】

例えば、GMM の場合、音響イベントごとに音響特徴量の各種別に対応する音響イベン

50

トモデルが得られる。例えば、音響特徴量列が F 種類 (F が 1 以上の整数) の音響特徴量 y (ただし、 $\{1, \dots, F\}$) からなる列 y_1, \dots, y_F である場合、各音響イベントに対応する音響イベントモデルは、それぞれ、以下のような確率モデル $p(y)$ を要素とする列 $p(y_1), \dots, p(y_F)$ となる。

【数 10】

$$p(y_i) = \sum_{j=1}^J \pi_j N(y_i | \mu_j, \Sigma_j) \quad (13)$$

ただし、 y は音響特徴量列 (ベクトル) の要素、 J は正規分布の混合数、 π_j は混合係数、 $N(\cdot)$ は正規分布の確率密度関数、 μ_j は分布の平均、 Σ_j は分布の分散である。 10

【0043】

或いは、音響イベントごとに音響特徴量列が対応付けられたものが音響イベントモデルとされてもよい。

【0044】

特徴量算出部 111 から出力された音響特徴量列は音響イベント判定部 112 に入力される。音響イベント判定部 112 は、入力された音響特徴量列と、音響イベントモデル DB113 に記憶されている複数の音響イベントモデルとをそれぞれ比較し、各フレーム (各要素番号 i に対応) の音響特徴量列に対応する音響イベントを決定する。例えば GMM が音響イベントモデルとして用いられる場合、音響イベント判定部 112 は、フレーム (要素番号 i) ごとに、入力された音響特徴量列の各要素 (ただし、 $\{1, \dots, F\}$) を各音響イベントに対応する式 (13) の各確率モデルに代入し、各音響イベントに対応する確率 $p(y_1) \times \dots \times p(y_F)$ を最大にする音響イベントを決定する。或いは、例えば音響イベントごとに音響特徴量列が対応付けられた音響イベントモデルの場合、音響イベント判定部 112 は、フレーム (要素番号 i) ごとに、入力された音響特徴量列との距離 (ユークリッド距離やコサイン距離) が最も近い音響イベントモデルに対応する音響イベントを選択する。 20

【0045】

音響イベント判定部 112 は、各要素番号 i に対して決定した音響イベントを表す音響イベントラベルを、行動ラベル候補付き音響信号列 15-s の各要素番号 i の要素に付与する。音響イベント判定部 112 は、この処理を入力された行動ラベル候補付き音響信号列 15-1, ..., 15-S のすべての要素 (すべての要素番号 i) について行い、その結果得られる音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, ..., 11-S を出力する。 30

【0046】

音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, ..., 11-S は、音響信号列合成部 101 に入力される。以降の処理は実施例 (1)-1-1 と同じである。

【0047】

なお、音響信号列合成部 101 で音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, ..., 11-S の合成処理を行うことに代えて、特徴量算出部 111 の前段で行動ラベル候補付き音響信号列 15-1, ..., 15-S の合成処理を行っても良いし、音響イベント判定部 112 の前段で音響特徴量列の合成処理を行っても良い。 40

【0048】

<実施例 (1)-1-3 : 図 4 >

[行動ラベル候補付き音響特徴量から行動 - 音響イベント生成モデルおよび行動遷移モデルを算出し、行動を分析]

本実施例では、行動の時間変化を考慮に入れ、行動ラベル候補付き音響特徴量から、行動 - 音響イベントの生成モデルを算出する。同時に、行動の時間遷移を表す行動遷移モデルも算出してもよく、行動の分析も行ってもよい。すなわち本実施例では、音響信号に対

応する音響特徴量を用いて音響イベントを決定し、それによって音響イベントラベル列を得てから学習処理を行い、行動・音響イベント生成モデルを生成する。なお、行動の時間遷移を表す行動遷移モデルも算出してもよく、行動の分析も行ってもよい。

【0049】

図4に例示するように、本実施例のモデル処理装置120は、音響イベント判定部112、音響イベントモデルデータベース(DB)113、音響信号列合成部101、モデル化部102、および記憶部103を有する。モデル処理装置120は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0050】

まず音響イベント判定部112に、行動ラベル候補付き音響特徴量列 $16 - 1, \dots, 16 - S$ が入力される。各行動ラベル候補付き音響特徴量列 $16 - s$ (ただし、 $s \in \{1, \dots, S\}$)は、時間区間ごとに区分された時系列の音響信号列、時間区間ごとに区分された音響信号列の各要素に対応する要素番号、長時間(数sec以上)ごとに決定されて時間区間ごとに付与された行動ラベル(「行動情報」に相当)、および音響信号列の時間区間ごとの音響特徴量列を含む。音響特徴量列の具体例は、実施例(1)-1-2で説明した通りである。

【0051】

音響イベント判定部112は、入力された行動ラベル候補付き音響特徴量列 $16 - s$ の音響特徴量列と、音響イベントモデルDB113に記憶されている複数の音響イベントモデルを、実施例(1)-1-2で説明したようにそれぞれ比較し、各フレーム(各要素番号*i*に対応)の音響特徴量列に対応する音響イベントを決定する。音響イベント判定部112は、各要素番号*i*に対して決定した音響イベントを表す音響イベントラベルを、行動ラベル候補付き音響特徴量列 $16 - s$ の各要素番号*i*の要素に付与する。音響イベント判定部112は、この処理を行動ラベル候補付き音響特徴量列 $16 - 1, \dots, 16 - S$ のすべての要素(すべての要素番号*i*)について行い、その結果得られる音響イベントラベルおよび行動ラベル候補付き音響信号列 $11 - 1, \dots, 11 - S$ を出力する。

【0052】

音響イベントラベルおよび行動ラベル候補付き音響信号列 $11 - 1, \dots, 11 - S$ は、音響信号列合成部101に入力される。以降の処理は実施例(1)-1-1,(1)-1-2と同じである。

【0053】

なお、音響信号列合成部101で音響イベントラベルおよび行動ラベル候補付き音響信号列 $11 - 1, \dots, 11 - S$ の合成処理を行うことに代えて、音響イベント判定部112の前段で音響特徴量列の合成処理を行っても良い。

【0054】

<実施例(1)-2-1:図5>

[音響イベントラベル付き音響信号列から行動を推定]

実施例(1)-2-1では、実施例(1)-1-1~(1)-1-3で説明したように得られた行動・音響イベント生成モデル14および/または行動遷移モデル17を用い、新たに入力された音響イベントラベル付き音響信号列から行動を推定する。

【0055】

図5に例示するように、本実施例のモデル処理装置200は、記憶部103および生成モデル比較部201を有する。モデル処理装置200は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0056】

まず生成モデル比較部201に音響イベントラベル付き音響信号列21が入力される。音響イベントラベル付き音響信号列21は、前述の時間区間ごとに区分された時系列の音響信号列、当該時間区間ごとに区分された音響信号列の各要素に対応する要素番号、および当該時間区間ごとに決定されて付与された音響イベントラベルを含む。要素番号および音響イベントラベルは、音響信号列の要素ごとに付与される。

10

20

30

40

50

【0057】

生成モデル比較部201は、入力された音響イベントラベル付き音響信号列21と、記憶部103に格納された行動・音響イベント生成モデル14および/または行動遷移モデル17とを比較し、音響イベントラベル付き音響信号列21に対し、最も適切であると判断した行動、又は最も適切なものから順番に複数個の行動を決定し、それらを推定結果として出力する。以下に音響イベントラベル付き音響信号列21と各生成モデルとの比較方法を例示する。

【0058】

[比較方法の例(1)-1]

音響イベントラベル付き音響信号列21が有する音響イベントの分布と、行動の候補に對応する確率P(音響イベント|行動の候補)で表される音響イベントの分布との近さに基づいて、行動が推定されてもよい。また、前時刻からの行動の候補の遷移確率に基づいて、行動が推定されてもよい。なお、行動の探索範囲は予め定められたものであってもよいし、音響イベントラベル付き音響信号列21がこの探索範囲を特定するための情報を含み、この情報によって探索範囲が定められてもよい。探索範囲を特定するための情報の例は、行動の候補を表す行動ラベル候補である。このような行動ラベル候補は、要素番号毎に設定されていてもよいし、複数の要素番号に対して共通であってもよい。この場合、探索範囲が適切に設定されていれば行動の推定精度が向上する。

10

【0059】

音響イベントラベル付き音響信号列21が有する音響イベントの分布と、行動の候補に對応する確率P(音響イベント|行動の候補)で表される音響イベントの分布との近さに基づいて、行動が推定される例では、まず生成モデル比較部201が、入力された音響イベントラベル付き音響信号列21から、以下のように $p(\cdot)$ (ただし、 $\{\cdot\}$, \dots , E)を算出する。

20

【数11】

$$p(\varepsilon) = \frac{C_\varepsilon + \gamma}{N_s' \times (1 + \gamma)} \quad (14)$$

ただし、 ε は事前に設定された緩和パラメータ(例えば0.01などの非負値)を表し、 C_ε は、音響イベントラベル付き音響信号列21で音響イベント ε を表す音響イベントラベルが付された要素の個数を表し、 N_s' は音響イベントラベル付き音響信号列21が含む音響信号列の要素数を表す。 $N_s' = N_s$ であってもよいし、 $N_s' > N_s$ であってもよい。

30

【0060】

次に生成モデル比較部201は、 $p(\cdot)$ と行動・音響イベント生成モデル14とを、下記に記すカルバックライブラー情報量(Kullback-Leibler divergence: KL divergence)やイエンセンシャノン情報量(Jensen-Shannon divergence: JS divergence)などの情報量基準に基づいて比較することで、入力された音響イベントラベル付き音響信号列21に対応する行動を推定する。

40

【数12】

$$(KL \text{ divergence}) = \sum_{\varepsilon=1}^E P(\varepsilon) \log \frac{P(\varepsilon)}{Q(\varepsilon)} \quad (15)$$

$$(JS \text{ divergence}) = \frac{1}{2} \left(\sum_{\varepsilon=1}^E P(\varepsilon) \log \frac{P(\varepsilon)}{\frac{P(\varepsilon) + Q(\varepsilon)}{2}} \right. \\ \left. + \sum_{\varepsilon=1}^E Q(\varepsilon) \log \frac{Q(\varepsilon)}{\frac{P(\varepsilon) + Q(\varepsilon)}{2}} \right) \quad (16)$$

10

【0061】

式(15)又は(16)の例の場合、生成モデル比較部201は、 $P(\cdot)$ に $p(\cdot)$ (ただし、 $\{1, \dots, E\}$)を代入し、 $Q(\cdot)$ に式(4)の $a(\cdot)$ (ただし、 $\{1, \dots, E\}$), $a(\{1, \dots, A\})$ を代入する。これにより、生成モデル比較部201は、各行動の候補 $a(\{1, \dots, A\})$ に対応する情報量(合計A個の情報量)を得る。行動の探索範囲が予め定められたものである場合にはAは定数である。一方、行動の探索範囲が音響イベントラベル付き音響信号列21に含まれる情報(行動ラベル候補等)によって特定される場合には、その探索範囲に属する行動の候補の総数がAとなる。

20

【0062】

生成モデル比較部201は、各行動 $a(\{1, \dots, A\})$ について算出された情報量のうち、最も小さな情報量に対応する行動、又は、最も小さな情報量から順番に選択した複数個の情報量に対応する複数個の行動を、音響イベントラベル付き音響信号列21に対応する行動として決定して出力する。

【0063】

生成モデル比較部201は、以下のように、行動遷移モデル17が表す遷移確率 $x_i^{x_{i-1}}$ と前時刻(要素番号*i*-1の時間区間)の行動推定結果 x_{i-1} を利用して、現時刻(要素番号*i*の時間区間)の行動 $x_i = a$ を推定しても良い。

30

【数13】

$$\arg \max_a P(x_i = a | x_{i-1}, \pi^{x_{i-1}} x_i) \quad (16-2)$$

【0064】

[比較方法の例(1)-2]

音響イベントラベル付き音響信号列21が有する音響イベントの時系列の並びに対する確率 P (音響イベント|行動の候補)のもとでの行動の候補の尤もらしさに基づいて、行動が推定されてもよい。また、音響イベントラベル付き音響信号列21が有する音響イベントの時系列の並びに対する確率 P (音響イベント|行動の候補)のもとでの行動の候補の尤もらしさおよび、前時刻からの行動の候補の遷移確率の両方にに基づいて行動が推定されても良い。ここでも上述のように、行動の探索範囲は予め定められたものであってもよいし、音響イベントラベル付き音響信号列21がこの探索範囲を特定するための情報(例えば、行動の候補を表す行動ラベル候補)を含み、この情報によって探索範囲が定められてもよい。

40

【0065】

以下に具体例を示す。以下の例では、生成モデル比較部201が、入力されたラベル付き音響信号列21に対し、行動-音響イベント生成モデル14のもとでの行動の尤度の和や積を求める。

50

【 0 0 6 6 】

行動 - 音響イベント生成モデル 1 4 のもとでの行動の尤度の和の例

【 数 1 4 】

$$P(x_i = a | e, \Omega, \beta)$$

$$= \sum_{i=1}^{N_s} P(x_i = a | e_i = \varepsilon, \Omega, \beta)$$

$$= \sum_{i=1}^{N_s} \frac{P(e_i | \Omega)}{\sum_{a=1}^A P(x_i = a, e_i = \varepsilon | \Omega, \beta)} \quad (17)$$

10

【 0 0 6 7 】

行動 - 音響イベント生成モデル 1 4 のもとでの行動の尤度の積の例

【 数 1 5 】

$$P(x_i = a | e, \Omega, \beta)$$

$$= \prod_{i=1}^{N_s} P(x_i = a | e_i = \varepsilon, \Omega, \beta)$$

$$= \prod_{i=1}^{N_s} \frac{P(e_i | \Omega)}{\sum_{a=1}^A P(x_i = a, e_i = \varepsilon | \Omega, \beta)} \quad (18)$$

20

【 0 0 6 8 】

30

ただし、式(17)および(18)の e_i は、入力された音響イベントラベル付き音響信号列 2 1 の要素番号 i に対応する音響イベントラベルが表す音響イベントを表す。式(17)および(18)は、式(4)の x_a と、入力された音響イベントラベル付き音響信号列 2 1 の e_i とから算出できる。

【 0 0 6 9 】

行動の探索範囲が予め定められたものである場合には、その探索範囲に属する各行動の候補 a について上記の尤度が算出され、行動の探索範囲が音響イベントラベル付き音響信号列 2 1 に含まれる情報（行動ラベル候補等）によって特定される場合には、その探索範囲に属する各行動の候補 a について上記の尤度が算出される。生成モデル比較部 2 0 1 は、各行動の候補 a について算出した尤度のうち、最も尤度の高い行動の候補、又は、最も尤度の高いものから順番に選択した複数個の行動の候補を、入力された音響イベントラベル付き音響信号列 2 1 に対応する行動として決定して出力する。

40

【 0 0 7 0 】

生成モデル比較部 2 0 1 は、以下のように、行動 - 音響イベント生成モデル 1 4 のもとでの行動の尤度の和や積および、行動遷移モデル 1 7 が表す遷移確率 $\times_{i-1} \times_i$ と前時刻（要素番号 $i - 1$ の時間区間）の行動推定結果 x_{i-1} を利用して、現時刻（要素番号 i の時間区間）の行動 $x_i = a$ を推定しても良い。

【 0 0 7 1 】

行動 - 音響イベント生成モデル 1 4 のもとでの行動の尤度の和および、行動遷移モデル 1 7 と前時刻の行動推定結果を利用する場合の例

50

【数16】

$$\begin{aligned}
 & P(x_i = a | e, \Omega, \beta, x_{i-1}, \pi^{x_{i-1} x_i}) \\
 &= \sum_{i=2}^{N_s} P(x_i = a | e_i = \varepsilon, \Omega, \beta, x_{i-1}, \pi^{x_{i-1} x_i}) \\
 &= \sum_{i=2}^{N_s} \frac{P(e_i | \Omega)}{\sum_{a=1}^A P(x_i = a, e_i = \varepsilon | \Omega, \beta)} \\
 &\quad \times P(x_i = a | x_{i-1}, \pi^{x_{i-1} x_i}) \quad (18-2)
 \end{aligned}$$

10

【0072】

行動 - 音響イベント生成モデル14のもとでの行動の尤度の積および、行動遷移モデル17と前時刻の行動推定結果を利用する場合の例

【数17】

$$\begin{aligned}
 & P(x_i = a | e, \Omega, \beta) \\
 &= \prod_{i=2}^{N_s} P(x_i = a | e_i = \varepsilon, \Omega, \beta, x_{i-1}, \pi^{x_{i-1} x_i}) \\
 &= \prod_{i=2}^{N_s} \frac{P(e_i | \Omega)}{\sum_{a=1}^A P(x_i = a, e_i = \varepsilon | \Omega, \beta)} \\
 &\quad \times P(x_i = a | x_{i-1}, \pi^{x_{i-1} x_i}) \quad (18-3)
 \end{aligned}$$

20

【0073】

30

<実施例(1)-2-2:図6>

[音響信号列から行動を推定]

本実施例では、実施例(1)-1-1～(1)-1-3で説明したように得られた行動 - 音響イベント生成モデル14および/または行動遷移モデル17を用い、新たに入力された音響信号列から行動を推定する。

【0074】

40

図6に例示するように、本実施例のモデル処理装置210は、特徴量算出部211、音響イベント判定部212、音響イベントモデルDB113、記憶部103、生成モデル比較部201を有する。モデル処理装置210は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0075】

まず特徴量算出部211に音響信号列22が入力される。音響信号列22は、時間区間に区分された時系列の音響信号列、および時間区間に区分された音響信号列の各要素に対応する要素番号を含む。

【0076】

特徴量算出部211は、音響信号列22から音響特徴量列(ベクトル)を算出して出力する。例えば特徴量算出部211は、前述した特徴量算出部111と同じ方法で音響特徴量列を算出する。

【0077】

音響イベント判定部212は、前述した音響イベント判定部112と同じ方法で、特徴

50

量算出部 211 から出力された音響特徴量列と、音響イベントモデル DB113 に記憶されている複数の音響イベントモデルとをそれぞれ比較し、音響信号列 22 の全ての要素について音響イベントを決定する。音響イベント判定部 212 は、決定した音響イベントを表す音響イベントラベルを音響信号列 22 の各要素に付することで、音響イベントラベル付き音響信号列 21 を生成して出力する。

【0078】

音響イベントラベル付き音響信号列 21 は、生成モデル比較部 201 に入力される。以降の処理は実施例(1)-2-1と同じである。

【0079】

< 実施例(1)-2-3 : 図 7 >

10

[音響特徴量から行動を推定]

本実施例では、実施例(1)-1-1～(1)-1-3で説明したように得られた行動 - 音響イベント生成モデル 14 および/または行動遷移モデル 17 を用い、新たに入力された音響特徴量列から行動を推定する。

【0080】

図 7 に例示するように、本実施例のモデル処理装置 220 は、音響イベント判定部 212、音響イベントモデル DB113、記憶部 103、生成モデル比較部 201 を有する。モデル処理装置 220 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0081】

20

まず音響イベント判定部 212 に音響特徴量列 23 が入力される。音響特徴量列 23 は、時間区間にごとに区分された時系列の音響信号列、当該時間区間にごとに区分された音響信号列の各要素に対応する要素番号、および音響信号列の音響特徴量列を含む。音響特徴量列の具体例は、実施例(1)-1-2 で説明した通りである。

【0082】

音響イベント判定部 212 は、前述した音響イベント判定部 112 と同じ方法で、音響特徴量列 23 の音響特徴量列と、音響イベントモデル DB113 に記憶されている複数の音響イベントモデルとをそれぞれ比較し、音響特徴量列 23 が含む音響信号列の全ての要素について音響イベントを決定する。音響イベント判定部 212 は、決定した各要素の音響イベントを表す音響イベントラベルを、音響特徴量列 23 が含む音響信号列に付することで、音響イベントラベル付き音響信号列 21 を生成して出力する。

30

【0083】

音響イベントラベル付き音響信号列 21 は、生成モデル比較部 201 に入力される。以降の処理は実施例(1)-2-1と同じである。

【0084】

< 実施例(1)-3-1 : 図 8 >

[音響イベントラベル付き音響信号列から行動を推定すると共に、音響イベントラベルおよび行動ラベル候補付き音響信号列から行動 - 音響イベント生成モデルおよび行動遷移モデルを算出]

本実施例は、実施例(1)-1-1 と実施例(1)-2-1との組み合わせである。本実施例では、音響イベントラベル付き音響信号列 21 を入力として行動を推定することに加え、音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, ..., 11-S を入力とし、行動 - 音響イベント生成モデル 14 の生成も行う。さらに行動遷移モデル 17 や行動ラベル 18 の生成を行ってもよい。さらに、推定結果である行動を表す行動ラベル候補を、音響イベントラベル付き音響信号列 21 が含む音響信号列の各要素に付与することで、音響イベントラベルおよび行動ラベル候補付き音響信号列を得、それを生成モデルの更新に利用することもできる。

40

【0085】

図 8 に例示するように、本実施例のモデル処理装置 300 は、記憶部 103, 303、音響信号列合成部 101、モデル化部 102、および生成モデル比較部 201 を有する。

50

モデル処理装置 300 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0086】

記憶部 303 には、前述した音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, . . . , 11-S、および前述した音響イベントラベル付き音響信号列 21 が格納されている。

【0087】

音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, . . . , 11-S は音響信号列合成部 101 に入力される。音響信号列合成部 101 およびモデル化部 102 は、実施例(1)-1-1 で説明したように、音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, . . . , 11-S から行動 - 音響イベント生成モデル 14 を生成し、それを記憶部 103 に格納する。また、実施例(1)-1-1 で説明したように、モデル化部 102 が行動遷移モデル 17 を生成して記憶部 103 に格納してもよいし、行動ラベル 18 を生成して記憶部 103 に格納してもよい。

10

【0088】

音響イベントラベル付き音響信号列 21 は生成モデル比較部 201 に入力される。音響イベントラベル付き音響信号列 21 は、実施例(1)-2-1 で説明したように、入力された音響イベントラベル付き音響信号列 21 と、記憶部 103 に格納された行動 - 音響イベント生成モデル 14 および/または行動遷移モデル 17 とを比較し、音響イベントラベル付き音響信号列 21 に対し、最も適切であると判断した行動、又は最も適切なものから順番に複数個の行動を選択し、それらを推定結果として出力する。

20

【0089】

さらに生成モデル比較部 201 は、決定した行動を表す行動ラベル候補を、音響イベントラベル付き音響信号列 21 の対応する各要素に付与し、それによって新たな音響イベントラベルおよび行動ラベル候補付き音響信号列 21' を生成して出力してもよい。この場合、音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, . . . , 11-S とともに音響信号列合成部 101 に入力されてもよい。この場合、音響信号列合成部 101 は、入力されたすべての音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, . . . , 11-S, 21' を時系列方向につなぎ合わせ(合成)、それによって一つの音響イベントラベルおよび行動ラベル候補付き音響信号列 11' (「第 3 音響イベントラベル列」および「第 3 行動ラベル候補」)を得て出力する。モデル化部 102 は、実施例(1)-1-1 のラベル付き音響信号列 11 (「音響イベントラベル列」および「行動ラベル候補」)に代えて、音響イベントラベルおよび行動ラベル候補付き音響信号列 11' (「第 3 音響イベントラベル列」および「第 3 行動ラベル候補」)を用い、再度、実施例(1)-1-1 で説明したように、行動 - 音響イベント生成モデル 14 を生成し、それを記憶部 103 に格納する。さらにモデル化部 102 が、ラベル付き音響信号列 11 に代えて、入力された音響イベントラベルおよび行動ラベル候補付き音響信号列 11' を用い、再度、実施例(1)-1-1 で説明したように、行動遷移モデル 17 を生成して記憶部 103 に格納してもよいし、行動ラベル 18 を生成して記憶部 103 に格納してもよい。

30

【0090】

また、生成モデル比較部 201 の処理およびモデル化部 102 の処理のどちらを先に行つても良い。ただし、モデル化部 102 の処理を行う前に生成モデル比較部 201 の処理を行う場合、記憶部 103 に予め得られた行動 - 音響イベント生成モデル 14 および/または行動遷移モデル 17 が格納されている必要がある。

40

【0091】

また、音響イベントラベルおよび行動ラベル候補付き音響信号列 21' が、新たに入力された音響イベントラベルおよび行動ラベル候補付き音響信号列とともに音響信号列合成部 101 に入力されてもよい。音響信号列合成部 101 は、これらを時系列方向につなぎ合わせ(合成)、それによって一つの音響イベントラベルおよび行動ラベル候補付き音響

50

信号列（「第3音響イベントラベル列」および「第3行動ラベル候補」）を得て、モデル化部102に送出し、モデル化部102が上述の処理を行ってもよい。

その他の処理は実施例(1)-1-1および実施例(1)-2-1と同じである。

【0092】

<実施例(1)-3-2：図9>

[音響信号列から行動を推定すると共に、行動ラベル候補付き音響信号列から行動-音響イベント生成モデルおよび行動遷移モデルを算出]

本実施例は実施例(1)-1-2と実施例(1)-2-2の組み合わせである。

本実施例では、行動ラベル候補付き音響信号列15-1, ..., 15-Sを入力として、学習によって、行動-音響イベント生成モデル14を算出する。さらに行行動遷移モデル17や行動ラベル18の生成を行ってもよい。さらに本実施例では、行動-音響イベント生成モデル14および/または行動遷移モデル17を用い、新たに入力された音響信号列22から行動を推定する。
10

【0093】

図9に例示するように、本実施例のモデル処理装置310は、特微量算出部111-1, ..., 111-S, 211、音響イベント判定部112-1, ..., 112-S, 212、音響イベントモデルDB113、および実施例(1)-3-1のモデル処理装置300(図8参照)を有する。

【0094】

行動ラベル候補付き音響信号列15-1, ..., 15-Sは、それぞれ特微量算出部111-1, ..., 111-Sに入力される。特微量算出部111-1, ..., 111-Sは、実施例(1)-1-2で説明したように、行動ラベル候補付き音響信号列15-1, ..., 15-Sから、それぞれ音響特微量列を得て出力する。音響イベント判定部112-1, ..., 112-Sは、それぞれ、前述の音響イベント判定部112と同様に、入力された音響特微量列と音響イベントモデルDB113に記憶されている複数の音響イベントモデルとから、音響イベントラベルおよび行動ラベル候補付き音響信号列11-1, ..., 11-Sを生成して出力する。
20

【0095】

音響信号列22は特微量算出部211に入力される。特微量算出部211は、実施例(1)-2-2で説明したように、音響信号列22から音響特微量列(ベクトル)を算出して出力する。音響イベント判定部212は、実施例(1)-2-2で説明したように、入力された音響特微量列と音響イベントモデルDB113に記憶されている複数の音響イベントモデルとから、音響イベントラベル付き音響信号列21を生成して出力する。
30

【0096】

音響イベントラベルおよび行動ラベル候補付き音響信号列11-1, ..., 11-S、音響イベントラベル付き音響信号列21は、記憶部303(図8)に格納される。以降の処理は実施例(1)-3-1と同じである。

【0097】

<実施例(1)-3-3：図10>

[音響特微量から行動を推定すると共に、行動ラベル候補付き音響信号列から行動-音響イベント生成モデルおよび行動遷移モデルを算出]

本実施例は実施例(1)-1-3と実施例(1)-2-3との組み合わせである。

本実施例では、行動ラベル候補付き音響特微量列16-1, ..., 16-Sを入力として、学習によって、行動-音響イベント生成モデル14を算出する。行動遷移モデル17や行動ラベル18の生成を行ってもよい。さらに本実施例では、行動-音響イベント生成モデル14を用い、新たに入力された音響特微量列23から行動を推定する。
40

【0098】

図10に例示するように、本実施例のモデル処理装置320は、音響イベント判定部112-1, ..., 112-S, 212、音響イベントモデルDB113、および前述のモデル処理装置300(図8参照)を有する。

【0099】

行動ラベル候補付き音響特徴量列 16 - 1, . . . , 16 - S は、それぞれ音響イベント判定部 112 - 1, . . . , 112 - S に入力される。音響イベント判定部 112 - 1, . . . , 112 - S は、それぞれ、前述の音響イベント判定部 112 と同様に、行動ラベル候補付き音響特徴量列 16 - 1, . . . , 16 - S の音響特徴量列と音響イベントモデル DB113 に記憶されている複数の音響イベントモデルとから、音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, . . . , 11 - S を生成して出力する。

【0100】

音響特徴量列 23 は音響イベント判定部 212 に入力される。音響イベント判定部 212 は、実施例(1) - 2 - 3 と同様に、入力された音響特徴量列 23 の音響特徴量列と音響イベントモデル DB113 に記憶されている複数の音響イベントモデルとから、音響イベントラベル付き音響信号列 21 を生成して出力する。10

【0101】

音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, . . . , 11 - S 、音響イベントラベル付き音響信号列 21 は、記憶部 303 (図 8) に格納される。以降の処理は実施例(1) - 3 - 1 と同じである。

【0102】

<実施例(1) - 4 - 1 : 図 11 >

[音響イベントラベル付き音響信号列から行動を推定すると共に、音響イベントラベルおよび行動ラベル候補付き音響信号列から行動 - 音響イベント生成モデルおよび行動遷移モデルを算出。特別なシンボルを用いる]20

本実施例は実施例(1) - 3 - 1 の変形である。

本実施例では、音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, . . . , 11 - S 、および音響イベントラベル付き音響信号列 21 を入力とし、行動 - 音響イベント生成モデル 14 を算出する。行動遷移モデル 17 や行動ラベル 18 の生成を行ってもよい。さらに、音響イベントラベル付き音響信号列 21 に対応する行動を推定する。さらに、推定された行動を表す行動ラベル候補を、音響イベントラベル付き音響信号列 21 が含む音響信号列の各要素に付与することで、音響イベントラベルおよび行動ラベル候補付き音響信号列を得、それを生成モデルの更新に利用することもできる。

【0103】

図 11 に例示するように、本実施例のモデル処理装置 400 は、記憶部 103、303 、音響信号列合成部 401 、モデル化部 402 、および生成モデル比較部 403 を有する。モデル処理装置 400 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。30

【0104】

音響信号列合成部 401 に、音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, . . . , 11 - S および音響イベントラベル付き音響信号列 21 が入力される。音響信号列合成部 401 は、音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, . . . , 11 - S 、音響イベントラベル付き音響信号列 21 を時系列方向につなぎ合わせ、それによって 1 つの音響イベントラベルおよび行動ラベル候補付き音響信号列 41 (以下、単に「ラベル付き音響信号列 41」という) を得て出力する。ここで前述のように、音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, . . . , 11 - S は、行動ラベル候補によって行動の候補が表された音響イベント (第 4 音響イベント) の時系列の並びを含む。一方、音響イベントラベル付き音響信号列 21 は、行動ラベル候補によって行動の候補が表されていない音響イベント (第 5 音響イベント) の時系列の並びを含む。したがって、ラベル付き音響信号列 41 が含む音響イベントの時系列の並びは、行動ラベル候補によって行動の候補が表された第 4 音響イベントの時系列の並び、および行動ラベル候補によって行動の候補が表されていない第 5 音響イベントの時系列の並びを含む。ラベル付き音響信号列 41 は、モデル化部 402 に入力される。なお、予め、音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, . . . , 11 -40

S、音響イベントラベル付き音響信号列21からラベル付き音響信号列41が得られている場合には、音響信号列合成部401を経由することなく、ラベル付き音響信号列41がそのままモデル化部402に入力されてもよい。

【0105】

モデル化部402は、実施例(1)-1-1のラベル付き音響信号列11に代えて、入力されたラベル付き音響信号列41を用い、実施例(1)-1-1のモデル化部102と同じ方法で、行動・音響イベント生成モデル14を算出する。行動遷移モデル17や行動ラベル18の生成を行ってもよい。ただし、モデル化部402は、上述の「第5音響イベント」を生じさせた行動の候補として特別なシンボルを用いて学習処理を行う。すなわちモデル化部402は、この特別なシンボルを一つの独立した特別な行動ラベル候補とみなす。この特別な行動ラベル候補は、一般的な何れの行動(料理、掃除など)を表すものでもない。例えば、この特別な行動ラベル候補が、「行動の候補が未知であること」を表してもよい。10

【0106】

すなわちモデル化部402は、行動の候補が音響イベントを生成する確率P(音響イベント|行動の候補)、およびこの特別なシンボルが音響イベントを生成する確率P(音響イベント|特別なシンボル)を得、それらに対応する行動・音響イベント生成モデル14を得る。行動・音響イベント生成モデル14は、記憶部103に格納される。行動遷移モデル17や行動ラベル18の生成を行った場合には、それらも記憶部103に格納される。20

【0107】

生成モデル比較部403には、行動・音響イベント生成モデル14および/または行動遷移モデル17が入力される。生成モデル比較部403は、入力された行動・音響イベント生成モデル14に対応する、確率P(音響イベント|行動の候補)と確率P(音響イベント|特別なシンボル)とを比較し、特別なシンボルに対応する行動の内容を推定し、それらの推定結果を出力する。また、生成モデル比較部403は、入力された行動遷移モデル17に対応する、遷移確率P(現時刻の行動の候補|前時刻の行動の候補)と遷移確率P(特別なシンボル|前時刻の行動の候補)とを比較し、特別なシンボルに対応する行動の内容を推定し、それらの推定結果を出力してもよい。例えば、生成モデル比較部403は、音響イベントラベル付き音響信号列21に対し、最も適切であると判断した行動、又は最も適切なものから順番に複数個の行動を選択し、それらを推定結果として出力する。30

【0108】

[比較方法の例示]

生成モデル比較部403による比較は、例えば、実施例(1)-2-1で説明した情報量基準に基づいて行うことができる。例えば、式(15)又は(16)の情報量基準を用い、確率P(音響イベント|行動の候補)と確率P(音響イベント|特別なシンボル)とを比較する場合、生成モデル比較部403は、P()に式(4)の_aのうち特別なシンボル_{a'} {1, ..., A}に対応する_{a'}(ただし、{1, ..., E})を代入し、Q()に他の行動の候補_{a''} {1, ..., A}に対応する_{a''}(ただし、{1, ..., E})を代入する。遷移確率P(現時刻の行動の候補|前時刻の行動の候補)と遷移確率P(特別なシンボル|前時刻の行動の候補)とを比較する場合、生成モデル比較部403は、P()に特別なシンボル_{a'}に対応するP($x_i = a' | x_{i-1}, \dots, x_{i-1}, x_i$)を代入し、Q()に他の行動の候補_{a''} {1, ..., A}に対応するP($x_i = a'' | x_{i-1}, \dots, x_{i-1}, x_i$)を代入する。40

【0109】

これにより、生成モデル比較部403は、他の行動_{a''} {1, ..., A}に対応する情報量(合計A-1個の情報量)を得る。生成モデル比較部403は、各行動_{a''}について算出された情報量のうち、最も小さな情報量に対応する行動の候補、又は、最も小さな情報量から順番に選択した複数個の情報量に対応する複50

数個の行動の候補を、音響イベントラベル付き音響信号列 2 1 に対応する行動として決定して出力する。

【 0 1 1 0 】

生成モデル比較部 4 0 3 は、以上のように決定した行動を表す行動ラベル候補を、音響イベントラベル付き音響信号列 2 1 の対応する各要素に付与し、それによって新たな音響イベントラベルおよび行動ラベル候補付き音響信号列 4 1 ' を生成して出力してもよい。この場合、音響イベントラベルおよび行動ラベル候補付き音響信号列 4 1 ' が、音響イベントラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S とともに音響信号列合成部 4 0 1 に入力されてもよい。この場合、音響信号列合成部 4 0 1 は、入力されたすべての音響イベントラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S , 4 1 ' を時系列方向につなぎ合わせ（合成）、それによって一つの音響イベントラベルおよび行動ラベル候補付き音響信号列 4 1 '' を得て出力する。モデル化部 4 0 2 は、実施例 (1) - 1 - 1 のラベル付き音響信号列 1 1 に代えて、音響イベントラベルおよび行動ラベル候補付き音響信号列 4 1 '' を用い、再度、実施例 (1) - 1 - 1 で説明したように、行動 - 音響イベント生成モデル 1 4 を生成し、それを記憶部 1 0 3 に格納する。さらにモデル化部 4 0 2 が、入力された音響イベントラベルおよび行動ラベル候補付き音響信号列 4 1 '' を用い、再度、実施例 (1) - 1 - 1 で説明したように、行動遷移モデル 1 7 を生成して記憶部 1 0 3 に格納してもよいし、行動ラベル 1 8 を生成して記憶部 1 0 3 に格納してもよい。

【 0 1 1 1 】

また、音響イベントラベルおよび行動ラベル候補付き音響信号列 2 1 ' が、新たに入力された音響イベントラベルおよび行動ラベル候補付き音響信号列とともに音響信号列合成部 4 0 1 に入力されてもよい。音響信号列合成部 4 0 1 は、これらを時系列方向につなぎ合わせ（合成）、それによって一つの音響イベントラベルおよび行動ラベル候補付き音響信号列を得て、モデル化部 4 0 2 に送出し、モデル化部 1 0 2 が上述の処理を行ってもよい。

その他の処理は実施例 (1) - 1 - 1 、実施例 (1) - 2 - 1 、および実施例 (1) - 3 - 1 と同じである。

【 0 1 1 2 】

< 実施例 (1) - 4 - 2 : 図 9 >

[音響信号列から行動を推定すると共に、行動ラベル候補付き音響信号列から行動 - 音響イベント生成モデルおよび行動遷移モデルを算出。特別なシンボルを用いる]

本実施例は実施例 (1) - 3 - 2 の変形である。

本実施例では、行動ラベル候補付き音響信号列 1 5 - 1 , . . . , 1 5 - S および音響信号列 2 2 を入力として、学習によって、行動 - 音響イベント生成モデル 1 4 を算出する。さらに行動遷移モデル 1 7 や行動ラベル 1 8 の生成を行ってもよい。さらに本実施例では、行動 - 音響イベント生成モデル 1 4 を用い、音響信号列 2 2 に対応する行動を推定する。

【 0 1 1 3 】

図 9 に例示するように、本実施例のモデル処理装置 4 1 0 は、特微量算出部 1 1 1 - 1 , . . . , 1 1 1 - S , 2 1 1 、音響イベント判定部 1 1 2 - 1 , . . . , 1 1 2 - S , 2 1 2 、音響イベントモデル DB 1 1 3 、および実施例 (1) - 4 - 1 のモデル処理装置 4 0 0 (図 1 1 参照) を有する。

【 0 1 1 4 】

行動ラベル候補付き音響信号列 1 5 - 1 , . . . , 1 5 - S は、それぞれ特微量算出部 1 1 1 - 1 , . . . , 1 1 1 - S に入力される。特微量算出部 1 1 1 - 1 , . . . , 1 1 1 - S は、実施例 (1) - 1 - 2 で説明したように、行動ラベル候補付き音響信号列 1 5 - 1 , . . . , 1 5 - S から、それぞれ音響特微量列を得て出力する。音響イベント判定部 1 1 2 - 1 , . . . , 1 1 2 - S は、それぞれ、前述の音響イベント判定部 1 1 2 と同様に、入力された音響特微量列と音響イベントモデル DB 1 1 3 に記憶されている複数の

10

20

30

40

50

音響イベントモデルとから、音響イベントラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S を生成して出力する。

【 0 1 1 5 】

音響信号列 2 2 は特徴量算出部 2 1 1 に入力される。特徴量算出部 2 1 1 および音響イベント判定部 2 1 2 は、実施例 (1) - 2 - 2 で説明したように、音響信号列 2 2 から音響特徴量列 (ベクトル) を算出して出力する。音響イベント判定部 2 1 2 は、実施例 (1) - 2 - 2 で説明したように、入力された音響特徴量列と音響イベントモデル DB 1 1 3 に記憶されている複数の音響イベントモデルとから、音響イベントラベル付き音響信号列 2 1 を生成して出力する。

【 0 1 1 6 】

音響イベントラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S 、音響イベントラベル付き音響信号列 2 1 は、記憶部 3 0 3 (図 1 1) に格納される。以降の処理は実施例 (1) - 3 - 1 と同じである。

【 0 1 1 7 】

< 実施例 (1) - 4 - 3 : 図 1 0 >

[音響特徴量から行動を推定すると共に、行動ラベル候補付き音響信号列から行動 - 音響イベント生成モデルおよび行動遷移モデルを算出。特別なシンボルを用いる]

本実施例は実施例 (1) - 3 - 3 の変形である。

本実施例では、行動ラベル候補付き音響特徴量列 1 6 - 1 , . . . , 1 6 - S および音響特徴量列 2 3 を入力として、学習によって、行動 - 音響イベント生成モデル 1 4 を算出する。行動遷移モデル 1 7 や行動ラベル 1 8 の生成を行ってもよい。さらに本実施例では、行動 - 音響イベント生成モデル 1 4 を用い、音響特徴量列 2 3 に対応する行動を推定する。

【 0 1 1 8 】

図 1 0 に例示するように、本実施例のモデル処理装置 4 2 0 は、音響イベント判定部 1 1 2 - 1 , . . . , 1 1 2 - S , 2 1 2 、音響イベントモデル DB 1 1 3 、および前述のモデル処理装置 4 0 0 (図 1 1 参照) を有する。

【 0 1 1 9 】

行動ラベル候補付き音響特徴量列 1 6 - 1 , . . . , 1 6 - S は、それぞれ音響イベント判定部 1 1 2 - 1 , . . . , 1 1 2 - S に入力される。音響イベント判定部 1 1 2 - 1 , . . . , 1 1 2 - S は、それぞれ、前述の音響イベント判定部 1 1 2 と同様に、入力された音響特徴量列と音響イベントモデル DB 1 1 3 に記憶されている複数の音響イベントモデルとから、音響イベントラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S を生成して出力する。

【 0 1 2 0 】

音響特徴量列 2 3 は音響イベント判定部 2 1 2 に入力される。音響イベント判定部 2 1 2 は、実施例 (1) - 2 - 3 と同様に、入力された音響特徴量列 2 3 の音響特徴量列と音響イベントモデル DB 1 1 3 に記憶されている複数の音響イベントモデルとから、音響イベントラベル付き音響信号列 2 1 を生成して出力する。

【 0 1 2 1 】

音響イベントラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S 、音響イベントラベル付き音響信号列 2 1 は、記憶部 3 0 3 (図 1 1) に格納される。以降の処理は実施例 (1) - 4 - 1 と同じである。

【 0 1 2 2 】

< 実施例 (2) - 1 - 1 : 図 1 2 >

[行動の時間変化および/または状況の時間変化を考慮に入れ、音響イベントラベルおよび行動ラベル候補付き音響信号から行動 - 音響イベント生成モデル、状況 - 音響イベント生成モデル、行動 - 状況生成モデル、および行動遷移モデルを算出し、同時に、行動や状況を分析]

本実施例では、行動の時間変化および/または状況の時間変化を考慮に入れ、音響イベ

10

20

30

40

50

ントラベルおよび行動ラベル候補付き音響信号から、状況 - 音響イベント生成モデルを算出する。同時に、行動 - 状況生成モデルを算出してもよく、行動 - 音響イベント生成モデルを算出してもよく、行動の時間遷移を表す行動遷移モデルを算出してもよく、状況の時間遷移を表す状況遷移モデルを算出してもよく、行動や状況の分析も行ってもよい。

【0123】

すなわち、本実施例では、各時間区間の音響信号を生じさせた音響イベントの時系列の並びを表す音響イベントラベル列、音響イベントを生じさせた行動の候補を表す行動ラベル候補、および音響イベントを生じさせた状況の候補の総数を用い、「音響イベントの時系列の並びにおける、行動の候補に対する音響イベントの組み合わせと行動の候補の時系列の並びおよび/または状況の時系列の並びと行動の候補に対する状況の候補の組み合わせとの同時事後確率」を最大化するための処理を行う学習処理を行う。例えば、この学習処理は、行動の時間遷移/または状況の時間遷移に基づく行動または状況の出現確率および、行動または状況による音響イベントの出現確率に基づいて、音響イベントの時系列の並びにおける、行動の候補に対する音響イベントの組み合わせと行動の候補の時系列の並びと行動に対する状況の候補の組み合わせとの同時事後確率を最大化するための処理である。例えば、音響イベントの時系列の並びからなる音響イベントの組み合わせに対する尤度関数 P （音響イベントの組み合わせ | 行動の候補の時系列の並びおよび/または状況の時系列の並び、行動の候補に対する状況の候補の組み合わせ）を最大にする処理が行われる。それにより、状況の候補が音響イベントを生成する確率 P （音響イベント | 状況の候補）を「状況 - 音響イベント生成モデル」として得る。また、この学習処理によつて、行動の候補が状況の候補を生成する確率 P （状況の候補 | 行動の候補）を「行動 - 状況生成モデル」として得てもよいし、行動の候補が音響イベントを生成する確率 P （音響イベント | 行動の候補）を「行動 - 音響イベント生成モデル」として得てもよいし、行動の候補の時系列の遷移確率を「行動遷移モデル」として得てもよいし、状況の候補の時系列の遷移確率を「状況遷移モデル」として得てもよい。なお「状況遷移モデル」とは、状況の候補の時系列の遷移確率である。状況遷移モデルの例は、ラベル付き音響信号列 11 に付与されたすべての行動ラベル候補を対象とした遷移確率 $, z_{i-1} z_i$ を (z_{i-1}, z_i) 要素とする $T \times T$ 行列である。ただし、遷移確率 $, z_{i-1} z_i$ は、前時刻での状況の候補 z_{i-1} (要素番号 $i-1$ での状況の候補) から現時刻での状況の候補 z_i (要素番号 i での状況の候補) への遷移確率を表し、 $, z_{i-1} z_i$ の上付き「 z_{i-1} 」は z_{i-1} を表し、下付き「 z_i 」は z_i を表し、 T は状況の候補の総数（状況の候補の種類の個数）を表す。さらに、この学習処理の過程で何れかの行動の候補を選択し、各音響イベントがどの行動によって生成されたのかを分析してもよい。この場合には選択した行動の候補を表す行動ラベルを出力してもよい。またさらに、この学習処理の過程で何れかの状況の候補を選択し、各音響イベントがどの状況によって生成されたのかを分析してもよい。この場合には選択した状況の候補を表す状況ラベルを出力してもよい。なお「状況の候補の総数」は、音響イベントラベル列とともに入力される値であってもよいし、予め定められた定数であってもよい。

【0124】

図 12 に例示するように、本実施例のモデル処理装置 100 は、音響信号列合成部 101、モデル化部 1102、および記憶部 1103 を有する。モデル処理装置 100 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0125】

まず音響信号列合成部 101 に、音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, ..., 11-S (ただし、S は 1 以上の整数) が入力される。図 13 に例示するように、音響イベントラベルおよび行動ラベル候補付き音響信号列 11-s (ただし、s = 1, ..., S) は、前述の時間区間ごとに区分された時系列の音響信号列、当該時間区間ごとに区分された音響信号列の各要素に対応する要素番号 i (ただし、i = 1, ..., N_s、N_s は正整数)、当該時間区間ごとに決定されて付与された音響イベ

10

20

30

40

50

ントラベル、当該時間区間に付与された行動ラベル候補、および響イ ventラベルおよび行動ラベル候補付き音響信号列 1 1 - s に対応する音響信号列番号 s を含む。上記時間区間に複数個の行動ラベル候補が付与されてもよいし（例えば、図 1 3 の音響信号列番号 1 - 要素番号 1 の時間区間）、単一の行動ラベル候補が付与されてもよい（例えば、図 1 3 の音響信号列番号 2 - 要素番号 1 の時間区間）。

【 0 1 2 6 】

複数個の音響イ ventラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S (以下、単に「ラベル付き音響信号列 1 1 - 1 , . . . , 1 1 - S 」という) が音響信号列合成部 1 0 1 に入力された場合、音響信号列合成部 1 0 1 は、それらを時系列方向につなぎ合わせ、それによって 1 つの音響イ ventラベルおよび行動ラベル候補付き音響信号列 1 1 (以下、単に「ラベル付き音響信号列 1 1 」という) を得て出力する（合成処理）。音響信号列合成部 1 0 1 に 1 つの音響信号列 1 1 - 1 のみが入力された場合、音響信号列合成部 1 0 1 はそれをラベル付き音響信号列 1 1 として出力する。音響信号列合成部 1 0 1 から出力された音響イ ventラベルおよび行動ラベル候補付き音響信号列は、モデル化部 1 1 0 2 に入力される。なお、音響信号列合成部 1 0 1 を経由することなく、1 つラベル付き音響信号列 1 1 がそのままモデル化部 1 1 0 2 に入力されてもよい。

【 0 1 2 7 】

モデル化部 1 1 0 2 は、以下の手順に従って、入力されたラベル付き音響信号列 1 1 から、行動 - 状況生成モデル 1 2 、状況 - 音響イ vent生成モデル 1 3 、および行動 - 音響イ vent生成モデル 1 4 の少なくとも何れかを生成し、それらを記憶部 1 1 0 3 に格納する。モデル化部 1 0 2 は、行動遷移モデル 1 7 を生成してもよいし、状況遷移モデルを生成してもよいし、行動ラベル 1 8 を生成してもよいし、状況ラベル 1 9 を生成してもよい。生成された行動遷移モデル 1 7 や状況遷移モデルや行動ラベル 1 8 や状況ラベル 1 9 も記憶部 1 1 0 3 に格納される。

【 0 1 2 8 】

[行動から音響イ vent が生成される過程の論理的説明]

モデル化部 1 1 0 2 は、ラベル付き音響信号列 1 1 から、行動の候補が状況の候補を生成する確率 P (状況の候補 | 行動の候補) としたときの行動 - 状況生成モデル 1 2 と、状況の候補が音響イ vent を生成する確率 P (音響イ vent | 状況の候補) としたときの状況 - 音響イ vent 生成モデル 1 3 と、行動の候補が音響イ vent を生成する確率 P (音響イ vent | 行動の候補) としたときの行動 - 音響イ vent 生成モデル 1 4 との少なくとも何れかを算出する。つまり本実施例では、行動の候補が状況の候補の生成確率を規定し、状況の候補が音響イ vent の生成確率を規定すると考え（つまり、間接的に行動の候補が音響イ vent の生成確率を規定するとも考えられる）、これらの関係を各生成モデルとして記述する。また、行動の候補の時系列の遷移確率を算出し、行動遷移モデル 1 7 を得てもよい。

【 0 1 2 9 】

状況の候補における音響イ vent の生成確率 、ラベル付き音響信号列 1 1 が有する行動ラベル候補が表す行動の候補のすべてからなる集合 、行動ラベル候補が表す行動の候補における状況の候補の生成確率 、および行動ラベル候補が表す行動の候補の遷移確率 および / または状況の候補の遷移確率 が与えられた場合における、音響イ vent の列 e の生成確率 P (e | , , , ,) , P (e | , , , ,) , P (e | , , , ,) は以下の通りである。

10

20

30

40

【数18】

$$P(e|\Pi, \Theta, \Phi', \Omega) = \prod_{s=1}^S P(e'_s | \Pi, \Theta, \Phi', a_s) \quad (21-1)$$

$$P(e|\Pi', \Theta, \Phi', \Omega) = \prod_{s=1}^S P(e'_s | \Pi', \Theta, \Phi', a_s) \quad (21-2)$$

$$P(e|\Pi, \Pi', \Theta, \Phi', \Omega) = \prod_{s=1}^S P(e'_s | \Pi, \Pi', \Theta, \Phi', a_s) \quad (21-3)$$

10

ただし、Aは行動の候補の総数、Tは状況の候補の総数（状況の候補の種類の個数）、Eは音響イベントの種類の個数、eはラベル付き音響信号列11に与えられた音響イベントの列（ベクトル）、 a_s は各時間区間の行動の候補a（ただし、 $a_s \in \{1, \dots, A\}$ ）が次の時間区間で行動の候補 a' （ただし、 $a' \in \{1, \dots, A\}$ ）に遷移する確率を (a, a') 要素とする $A \times A$ 行列、 t は各時間区間の状況の候補t（ただし、 $t \in \{1, \dots, T\}$ ）が次の時間区間で行動の候補 t' （ただし、 $t' \in \{1, \dots, T\}$ ）に遷移する確率を (t, t') 要素とする $T \times T$ 行列、 a は行動の候補a（ただし、 $a \in \{1, \dots, A\}$ ）が状況の候補t（ただし、 $t \in \{1, \dots, T\}$ ）を生成する確率 $P(t|a)$ を (a, t) 要素とする $A \times T$ 行列、 e は状況の候補t（ただし、 $t \in \{1, \dots, T\}$ ）が音響イベント（ただし、 $e \in \{1, \dots, E\}$ ）を生成する確率 $P(e|t)$ を (t, e) 要素とする $T \times E$ 行列、 e' は各音響信号列がとり得る行動ラベル候補を音響信号列分並べて列とした、集合のベクトル、 e'_s は各ラベル付き音響信号列11-sに与えられた音響イベントの列（ N_s 次元ベクトル： N_s はラベル付き音響信号列11-sに対応する音響イベントの個数）、 a_s はラベル付き音響信号列11-sに付与され得る行動の集合、Sはラベル付き音響信号列11を構成するラベル付き音響信号列11-sの個数を表す。なお、 a と a_s の関係は以下のように表される。

$$= [a_1, a_2, \dots, a_s, \dots, a_S]$$

【0130】

行動の候補の集合 a_s 、行動の候補aの遷移確率 a の事前分布のパラメータ（Dirichlet分布に従うものとする）、行動の候補aにおける状況の候補tの生成確率 a の事前分布のパラメータ（Dirichlet分布に従うものとする）、状況の候補tにおける音響イベントの生成確率 t の事前分布のパラメータ（Dirichlet分布に従うものとする） e'_s が与えられたときの、音響イベントの列 e'_s の生成確率 $P(e'_s, \dots, e'_s | a_s, t)$ 、行動の候補の集合 a_s 、状況の候補tの遷移確率 t の事前分布のパラメータ（Dirichlet分布に従うものとする） a 、パラメータ a が与えられたときの、音響イベントの列 e'_s の生成確率 $P(e'_s, \dots, e'_s | a_s, a)$ 、および行動の候補の集合 a_s 、パラメータ a が与えられたときの、音響イベントの列 e'_s の生成確率 $P(e'_s, \dots, e'_s | a_s, a)$ は、以下の通りである。

20

30

40

【数19】

$$\begin{aligned}
 & P(e'_s, \Pi, \Theta, \Phi' | \gamma, \alpha, \beta', a_s) \\
 & = P(\pi_a | \gamma) P(\phi_a | \alpha) P(\phi_t | \beta') \\
 & \times \prod_{i=1}^{N_s} P(e_i | z_i, \phi_t) P(z_i | x_i, \theta_a) P(x_i | \pi_{x_i}^{x_{i-1}}, x_{i-1}, a_s) \quad (22-1)
 \end{aligned}$$

$$\begin{aligned}
 & P(e'_s, \Pi', \Theta, \Phi' | \gamma', \alpha, \beta', a_s) \\
 & = P(\pi'_t | \gamma') P(\phi_a | \alpha) P(\phi_t | \beta') \\
 & \times \prod_{i=1}^{N_s} P(e_i | z_i, \phi_t) P(z_i | x_i, \theta_a) P(z_i | \pi_{z_i}^{z_{i-1}}, z_{i-1}) \quad (22-2)
 \end{aligned}$$

$$\begin{aligned}
 & P(e'_s, \Pi, \Pi', \Theta, \Phi' | \gamma, \gamma', \alpha, \beta', a_s) \\
 & = P(\pi_a | \gamma) P(\pi'_t | \gamma') P(\phi_a | \alpha) P(\phi_t | \beta') \\
 & \times \prod_{i=1}^{N_s} P(e_i | z_i, \phi_t) P(z_i | x_i, \theta_a) P(x_i | \pi_{x_i}^{x_{i-1}}, x_{i-1}, a_s) \\
 & \times P(z_i | \pi_{z_i}^{z_{i-1}}, z_{i-1}) \quad (22-3)
 \end{aligned}$$

ただし、 e_i はラベル付き音響信号列 11 - s の要素番号 i に対応する音響イベント、 N_s はラベル付き音響信号列 11 - s に対応する要素数（要素番号 i の最大値）、 z_i はラベル付き音響信号列 11 - s の要素番号 i に対応する音響信号が表状況の候補、 x_i はラベル付き音響信号列 11 - s の要素番号 i に対応する行動ラベル候補が表す行動の候補、 $\pi_{x_i}^{x_{i-1}}$ はラベル付き音響信号列 11 - s の要素番号 i - 1 に対応する行動ラベル候補が表す行動の候補 x_{i-1} から要素番号 i に対応する行動ラベル候補が表す行動の候補 x_i への遷移確率を表す。また、 t は状況の候補 t が音響イベント（ただし、 $\{1, \dots, E\}$ ）を生成する確率 $P(\cdot | t)$ を番目の要素とする E 次元ベクトルを表す。 a は行動の候補 a が状況の候補 t（ただし、 $t \in \{1, \dots, T\}$ ）を生成する確率 $P(t | a)$ を t 番目の要素とする T 次元ベクトルを表す。

【0131】

[生成モデルの算出過程の説明]

モデル化部 1102 は、入力されたラベル付き音響信号列 11 から、状況の候補が音響イベントを生成する確率 $P(\text{音響イベント} | \text{状況の候補})$ を算出し、それを状況 - 音響イベント生成モデル 13 として記憶部 1103 に格納する。モデル化部 1102 は、行動の候補が状況の候補を生成する確率 $P(\text{状況の候補} | \text{行動の候補})$ を算出し、それを行動 - 状況生成モデル 12 として記憶部 1103 に格納してもよい。また、モデル化部 1102 は、行動の候補が音響イベントを生成する確率 $P(\text{音響イベント} | \text{行動の候補})$ を行動 - 音響イベント生成モデル 14 として算出してもよいし、行動の候補の遷移確率を算出し、それを行動遷移モデル 17 として記憶部 1103 に格納してもよい。

【0132】

さらにモデル化部 1102 は、入力されたラベル付き音響信号列 11 が有する行動ラベル候補から尤もらしいラベルを選択し、選択した行動ラベル候補からなる行動ラベル列 18 を外部に出力したり、記憶部 1103 に格納したりしてもよい。尤もらしいラベルは、

10

20

30

40

50

各時間区間にに対して1つでもよいし、複数個でもよい。また、各時間区間において尤もらしさが閾値を超えたすべての行動ラベル候補を選択し、選択した時間区間ごとに単一または複数の行動ラベル候補からなる行動ラベル18の列を外部に出力したり、記憶部1103に格納したりしてもよい。

【0133】

またモデル化部1102は、入力されたラベル付き音響信号列11に対して尤もらしい状況を表す状況ラベル19を選択して外部に出力したり、記憶部1103に格納したりしてもよい。選択される状況ラベル19は、各時間区間にに対して1個でもよいし、複数個でもよい。また、各時間区間において尤もらしさが閾値を超えたすべての状況を表すラベルを状況ラベル19の列を外部に出力したり、記憶部1103に格納したりしてもよい。

10

【0134】

行動 - 状況生成モデル12および状況 - 音響イベント生成モデル13の算出は、入力されたラベル付き音響信号列11の音響イベントの時系列の並びにおける、行動の候補に対する音響イベントの組み合わせと、行動の候補の時系列の並びと、行動の候補に対する状況の候補の組み合わせとの同時事後確率を最大化する処理を用いて行われる。例えば、マルコフ連鎖モンテカルロ法(MCMC法:Markov Chain Monte Carlo methods)や変分ベイズ法(VB法:Variational Bayes methods)などの手法を用いることができる。MCMC法には、M-Hアルゴリズムやギブスサンプリングなどの手法がある。ここでは一例としてギブスサンプリングによる生成モデルの算出手法を説明する。

【0135】

20

(i)モデル化部1102は、ハイパーパラメータとしてパラメータ α , β ならびにおよび/または γ を事前に決定する。

(ii)モデル化部1102は、全てのラベル付き音響信号列11-s(ただし $s = 1, \dots, S$)の全ての要素番号 i に対応する要素に行動の候補 x_i をランダムに割り当てる(割り当てはラベル付き音響信号列11-sに与えられた行動ラベル候補が表す行動の候補からランダムに行う)。あるいは、要素番号 i に対応する要素への行動の候補 x_i の割り当ては、事前にモデル化部1102に設定しておいた割り当てであってもよい。

【0136】

(iii)モデル化部1102は、全てのラベル付き音響信号列11-s(ただし $s = 1, \dots, S$)の各要素番号 i (要素番号 $i-1$ または $i+1$ に対応する要素が存在しない要素番号 i を除く)について、以下の(iii-a)および(iii-b)を交互に、規定の回数(正值であり、例えば1~3000回程度)、もしくは、所望の結果が得られるまで(例えば、割り当ての前後において、行動の割り当て先の変化が一定の閾値(例えば0.3%)以下になるまでなど)繰り返す。

30

【0137】

(iii-a)は以下の(iii-a-1)および(iii-a-2)からなる。

(iii-a-1)モデル化部1102は、以下の式(23a-1)~(23a-3)の何れかにより、更新対象のラベル付き音響信号列の要素番号 i を除いた要素番号に対応する音響信号列に対して、要素番号 i の要素に行動の候補 a に割り当てられる確率分布 $P(x_i = a | x_{-i}, e, \dots, z_{-i})$, $P(x_i = a | x_{-i}, z_{-i}, e, \dots, ')$, $P(x_i = a | x_{-i}, z_{-i}, e, \dots, ', ')$ の何れかを更新する。

40

【数20】

$$P(x_i = a | x_{-i}, e, \Omega, \alpha, \beta, \gamma)$$

$$\propto \frac{C_{ta}^{TA} + \alpha}{(\sum_{t=1}^T C_{ta}^{TA}) + T\alpha} \cdot \frac{C_{et}^{ET} + \beta}{(\sum_{e=1}^E C_{ea}^{ET}) + E\beta} \\ \times \frac{(R_{x_i}^{x_{i-1}} + \gamma)(R_{x_{i+1}}^{x_i} + I(x_{i-1} = x_i) \times I(x_i = x_{i+1}) + \gamma)}{(\sum_{x'_{i+1} \in \Omega} R_{x'_{i+1}}^{x_i}) + I(x_{i-1} = x_i) + A\gamma} \quad (23a-1)$$

10

$$P(x_i = a | x_{-i}, z_{-i}, e, \Omega, \alpha, \beta, \gamma')$$

$$\propto \frac{C_{ta}^{TA} + \alpha}{(\sum_{t=1}^T C_{ta}^{TA}) + T\alpha} \cdot \frac{C_{et}^{ET} + \beta}{(\sum_{e=1}^E C_{ea}^{ET}) + E\beta} \\ \times \frac{(R_{z_i}^{z_{i-1}} + \gamma')(R_{z_{i+1}}^{z_i} + I(z_{i-1} = z_i) \times I(z_i = z_{i+1}) + \gamma')}{(\sum_{z'_{i+1} \in \Omega'} R_{z'_{i+1}}^{z_i}) + I(z_{i-1} = z_i) + T\gamma'} \quad (23a-2)$$

20

$$P(x_i = a | x_{-i}, z_{-i}, e, \Omega, \alpha, \beta, \gamma, \gamma')$$

$$\propto \frac{C_{ta}^{TA} + \alpha}{(\sum_{t=1}^T C_{ta}^{TA}) + T\alpha} \cdot \frac{C_{et}^{ET} + \beta}{(\sum_{e=1}^E C_{ea}^{ET}) + E\beta} \\ \times \frac{(R_{x_i}^{x_{i-1}} + \gamma)(R_{x_{i+1}}^{x_i} + I(x_{i-1} = x_i) \times I(x_i = x_{i+1}) + \gamma)}{(\sum_{x'_{i+1} \in \Omega} R_{x'_{i+1}}^{x_i}) + I(x_{i-1} = x_i) + A\gamma} \\ \times \frac{(R_{z_i}^{z_{i-1}} + \gamma')(R_{z_{i+1}}^{z_i} + I(z_{i-1} = z_i) \times I(z_i = z_{i+1}) + \gamma')}{(\sum_{z'_{i+1} \in \Omega'} R_{z'_{i+1}}^{z_i}) + I(z_{i-1} = z_i) + T\gamma'} \quad (23a-3)$$

30

ただし、 C_{ta}^{TA} は更新対象のラベル付き音響信号列（初期値はラベル付き音響信号列11）で状況の候補 t に行動の候補 a が割り当てられた回数を表す。 z_{-i} は要素番号 i 以外の要素番号に対応する状況からなる列を表す。なお、表記制約上の都合から「 C_{ta}^{TA} 」と表記するが、本来は式(23a-1)～(23a-3)に示すように「 C_{ta}^{TA} 」の「 TA 」は「 ta 」の上に表記される。 $R_{z_i z_{i+1}}^{z_i}$ は、更新対象のラベル付き音響信号列に対応するすべて状況の候補の並びの中で、状況の候補 $z_i = t = \{1, \dots, T\}$ と同一の候補から行動の候補 z_{i+1} と同一の候補に遷移した回数を表す。その他は、前述の式(3)に用いた記号の定義と同じである。

40

【0138】

(iii-a-2)モデル化部1102は、上記の更新式(23a-1)～(23a-3)の何れかで得られた確率分布 $P(x_i = a | x_{-i}, e, \dots, \dots)$ 、 $P(x_i = a | x_{-i}, z_{-i}, e, \dots, \dots)$ 、 $P(x_i = a | x_{-i}, z_{-i}, e, \dots, \dots)$ の何れかに従って、各要素番号 i の要素に割り当てる行動の候補 x_i をサンプリング（選択）する。モデル化部1102は、このようにサンプリングした行動の候補 x_i を要素番号 i の要素に割り当て、要素番号 i に対応する行動の候補を更新する。これにより更新対象のラベル付き音響信号列を更新する。

50

【0139】

(iii-b-1) モデル化部 1102 は、以下の式により、更新対象のラベル付き音響信号列の要素番号 i を除いた要素番号に対応する音響信号列に対して、要素番号 i に対応する音響イベントが状況の候補 t に割り当てる確率分布 $P(z_i = t | z_{-i}, e, \alpha, \beta')$ を更新する。

【数21】

$$P(z_i = t | z_{-i}, e, \Omega, \alpha, \beta') \quad (23b)$$

$$\propto \frac{C_{et}^{ET} + \beta'}{\left(\sum_{e'=1}^E C_{e't}^{ET}\right) + E\beta'} \times \frac{C_{ta}^{TA} + \alpha}{\left(\sum_{t'=1}^T C_{t'a}^{TA}\right) + T\alpha} \quad (23b)$$

10

ただし、 C_{et}^{ET} は更新対象のラベル付き音響信号列で音響イベント e に状況の候補 t が割り当たされた回数を表す。その他は、前述の式(23a-1)～(23a-3)に用いた記号の定義と同じである。

【0140】

(iii-b-2) モデル化部 1102 は、上記の更新式(23b)で得られた確率分布 $P(z_i = t | z_{-i}, e, \alpha, \beta')$ に従って、各要素番号 i の要素に割り当てる状況の候補をサンプリング(選択)する。モデル化部 1102 は、このようにサンプリングした状況の候補を各要素番号 i の要素に割り当て、更新対象のラベル付き音響信号列を更新する。

20

【0141】

(iv) モデル化部 1102 は、上記の(iii-a)(iii-b)の繰り返しによって最終的に得られた C_{et}^{ET} , C_{ta}^{TA} ならびに $R^{x_i \times i+1}$ および / または $R^{z_i \times z_{i+1}}$ を用い、以下を計算する。なお、式(26-1)および(26-2)の両方が計算されてもよいし、その何れか一方のみが計算されてもよい。

【数22】

$$\phi_{et} = \frac{C_{et}^{ET} + \beta'}{\left(\sum_{e'=1}^E C_{e't}^{ET}\right) + E\beta'} \quad (24)$$

30

$$\theta_{ta} = \frac{C_{ta}^{TA} + \alpha}{\left(\sum_{t'=1}^T C_{t'a}^{TA}\right) + T\alpha} \quad (25)$$

$$\pi_{x_{i+1}}^{x_i} = \frac{(R_{x_i}^{x_{i-1}} + \gamma)(R_{x_{i+1}}^{x_i} + \gamma)}{\left\{ \left(\sum_{x'_i \in \Omega} R_{x'_i}^{x_{i-1}} \right) + A\gamma \right\} \left\{ \left(\sum_{x'_{i+1} \in \Omega} R_{x'_{i+1}}^{x_i} \right) + A\gamma \right\}} \quad (26-1)$$

$$\pi_{z_{i+1}}^{z_i} = \frac{(R_{z_i}^{z_{i-1}} + \gamma')(R_{z_{i+1}}^{z_i} + \gamma')}{\left\{ \left(\sum_{z'_i \in \Omega''} R_{z'_i}^{z_{i-1}} \right) + T\gamma' \right\} \left\{ \left(\sum_{z'_{i+1} \in \Omega''} R_{z'_{i+1}}^{z_i} \right) + T\gamma' \right\}} \quad (26-2)$$

40

【0142】

これによってモデル化部 1102 は、状況が音響イベントを生成する確率 ϕ_{et} (ただし、 $t \in \{1, \dots, T\}$, $e \in \{1, \dots, E\}$) の集合を得、それを状況 - 音響イベント生成モデル 13 とする。例えばモデル化部 1102 は、確率 ϕ_{et} を (t, e) 要素とする $T \times E$ 行列を状況 - 音響イベント生成モデル 13 とする。また、モデル化部 1102 は、行動が状況を生成する確率 θ_{ta} (ただし、 $a \in \{1, \dots, A\}$, $t \in \{1, \dots, T\}$) の集合を得、それを行動 - 状況生成モデル 12 とする。例えばモデル化部 1102 は、確率 θ_{ta} を (a, t) 要素とする $A \times T$ 行列を行動 - 状況生成モデル 12 とする。また、モデル化部 1102 は、これによって行動の候補の遷移確率 $\pi_{x_{i+1}}^{x_i}$

50

x_{i+1} (ただし、 $x_i, x_{i+1} \in \{1, \dots, A\}$) の集合を得、それを行動遷移モデル 17 としてもよい。例えばモデル化部 102 は、 $x_i, x_{i+1} \in \{1, \dots, A\}$ に対応する遷移確率 $x^i_{x_{i+1}}$ を (x_i, x_{i+1}) 要素とする $A \times A$ 行列を行動遷移モデル 17 とする。また、モデル化部 1102 は、これによって行動の候補の遷移確率 $x^i_{z_{i+1}}$ (ただし、 $z_i, z_{i+1} \in \{1, \dots, T\}$) の集合を得、それを状況遷移モデル 17 としてもよい。例えばモデル化部 102 は、 $z_i, z_{i+1} \in \{1, \dots, T\}$ に対応する遷移確率 $z^i_{z_{i+1}}$ を (z_i, z_{i+1}) 要素とする $T \times T$ 行列を状況遷移モデルとする。

【0143】

或いは、モデル化部 1102 は、(iii-b-1) の繰り返し処理時に、式 (23b) の算出過程で確率 t_a および確率 t_a をそれぞれ 1 個以上サンプリング (計算) し、式 (24) (25) に代えて、サンプリングされた確率 t_a の平均値および確率 t_a の平均値を用い、状況 - 音響イベント生成モデル 13 および行動 - 状況生成モデル 12 を得てもよい。同様に、モデル化部 1102 は、(iii-1) の繰り返し処理時に、式 (23a-1) ~ (23a-3) の算出過程で遷移確率 $x^i_{x_{i+1}}$ を 1 個以上サンプリング (計算) し、式 (26-1) に代えて、サンプリングされた遷移確率 $x^i_{x_{i+1}}$ の平均値を用い、行動遷移モデル 17 を得てもよい。同様に、モデル化部 1102 は、(iii-1) の繰り返し処理時に、式 (23a-1) ~ (23a-3) の算出過程で遷移確率 $z^i_{z_{i+1}}$ を 1 個以上サンプリング (計算) し、式 (26-2) に代えて、サンプリングされた遷移確率 $z^i_{z_{i+1}}$ の平均値を用い、状況遷移を得てもよい。10 20

【0144】

さらにモデル化部 1102 は、以下のように周辺化することで、行動の候補 a (ただし、 $a \in \{1, \dots, A\}$) がラベル付き音響信号列 11-s に与えられた音響イベントの列 e'_s を生成する確率 $P(e'_s | \dots)$ の集合を得、これらに対応する行動 - 音響イベント生成モデル 14 を得てもよい。

【数 23】

$$P(e'_s | \Theta, \Phi, \Omega) = \prod_{i=1}^{N_s} \sum_{t=1}^T \phi_{e'_s t} \theta_{ta_i} \quad (27)$$

30

ただし、 e'_s は要素番号 i に対応する音響イベント、 a_i は要素番号 i に対応する行動の候補を表す。

【0145】

また、上記更新過程でサンプリング (選択) し、各要素番号 i の要素に割り当てた行動の候補 x_i を分析することで、各ラベル付き音響信号列 11-s の音響信号列がどの行動によって生成されたかを知ることができる。モデル化部 1102 は、このように各要素番号 i の要素に割り当てた行動の候補 x_i を表すラベル、またはそれらの一部を行動ラベル 18 として出力してもよい。

【0146】

また、上記更新過程でサンプリング (選択) し、各要素番号 i の要素に割り当てた状況の候補 z_i を分析することで、各ラベル付き音響信号列 11-s の音響信号列がどの状況によって生成されたかを知ることができる。モデル化部 1102 は、このように各要素番号 i の要素に割り当てた状況の候補 z_i を表すラベル、またはそれらの一部を状況ラベル 19 として出力してもよい。40

【0147】

本実施例では、行動が音響イベントを生成する確率のみではなく、行動の時間遷移確率および / または状況の時間遷移確率も考慮して学習処理を行うことで、遷移確率によって特徴付けられる行動やそれに対応する状況を精度よくモデル化できる。また、遷移確率によって特徴付けられる行動やそれに対応する状況を精度よく分析できる。

【0148】

50

<実施例(2)-1-2:図14>

[行動ラベル候補付き音響信号列から行動・音響イベント生成モデル、状況・音響イベント生成モデル、行動・状況生成モデル、および行動遷移モデルを算出し、行動や状況を分析]

本実施例では、行動ラベル候補付き音響信号列 15 - 1, ..., 15 - S を入力とし、行動の時間変化および/または状況の時間変化を考慮に入れ、状況・音響イベント生成モデル 13 を算出する。すなわち、本実施例では、音響信号から音響特徴量を得、得られた音響特徴量を用いて音響イベントを決定し、それによって音響イベントラベル列を得てから学習処理を行い、状況・音響イベント生成モデル 13 を算出する。この際、行動・状況生成モデル 12 を算出してもよいし、状況の時間遷移を表す状況遷移モデルを算出してもよいし、行動・音響イベント生成モデル 14 を算出してもよいし、行動遷移モデル 17 を算出してもよいし、行動や状況の分析を行ってもよい。

【0149】

図14に例示するように、本実施例のモデル処理装置 1110 は、特徴量算出部 111、音響イベント判定部 112、音響イベントモデルデータベース(DB) 113、音響信号列合成部 101、モデル化部 1102、および記憶部 1103 を有する。モデル処理装置 1110 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0150】

まず特徴量算出部 111 に行動ラベル候補付き音響信号列 15 - 1, ..., 15 - S が入力される。特徴量算出部 111 は、各行動ラベル候補付き音響信号列 15 - s (ただし、s ∈ {1, ..., S}) が含む音響信号列から、各行動ラベル候補付き音響信号列 15 - s に対応する音響特徴量列 (ベクトル) を算出して出力する。音響特徴量列の具体例は、実施例(1)-1-2と同じである。

【0151】

音響イベント判定部 112 は、実施例(1)-1-2 で例示した方法等によって、特徴量算出部 111 で得られた音響特徴量を用いて各要素番号 i に対して音響イベントを決定する。音響イベント判定部 112 は、各要素番号 i に対して決定した音響イベントを表す音響イベントラベルを、行動ラベル候補付き音響信号列 15 - s の各要素番号 i の要素に付与する。音響イベント判定部 112 は、この処理を入力された行動ラベル候補付き音響信号列 15 - 1, ..., 15 - S のすべての要素 (すべての要素番号 i) について行い、その結果得られる音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, ..., 11 - S を出力する。

【0152】

音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, ..., 11 - S は、音響信号列合成部 101 に入力される。以降の処理は実施例(2)-1-1と同じである。

【0153】

なお、音響信号列合成部 101 で音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, ..., 11 - S の合成処理を行うことに代えて、特徴量算出部 111 の前段で行動ラベル候補付き音響信号列 15 - 1, ..., 15 - S の合成処理を行っても良いし、音響イベント判定部 112 の前段で音響特徴量列の合成処理を行っても良い。

【0154】

<実施例(2)-1-3:図15>

[行動ラベル候補付き音響特徴量から行動・音響イベント生成モデル、状況・音響イベント生成モデル、行動・状況生成モデル、および行動遷移モデルを算出し、行動や状況を分析]

本実施例では、行動ラベル候補付き音響特徴量列 16 - 1, ..., 16 - S を入力とし、行動の時間変化および/または状況の時間変化を考慮に入れ、状況・音響イベント生成モデル 13 を算出する。すなわち、本実施例では、入力された音響特徴量を用いて音響

10

20

30

40

50

イベントを決定し、それによって音響イベントラベル列を得てから学習処理を行い、状況 - 音響イベント生成モデル 13 を算出する。この際、行動 - 状況生成モデル 12 を算出してもよいし、行動 - 音響イベント生成モデル 14 を算出してもよいし、行動遷移モデル 17 を算出してもよいし、状況の時間遷移を表す状況遷移モデルを算出してもよいし、行動や状況の分析を行ってもよい。

【0155】

図 15 に例示するように、本実施例のモデル処理装置 1120 は、音響イベント判定部 112、音響イベントモデルデータベース (DB) 113、音響信号列合成部 101、モデル化部 1102、および記憶部 1103 を有する。モデル処理装置 1120 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。10

【0156】

まず音響イベント判定部 112 に、行動ラベル候補付き音響特徴量列 16-1, ..., 16-S が入力される。音響イベント判定部 112 は、実施例 (1)-1-2 で例示した方法等によって、音響特徴量列 16-1, ..., 16-S を用いて各要素番号 i に対して音響イベントを決定する。音響イベント判定部 112 は、各要素番号 i に対して決定した音響イベントを表す音響イベントラベルを、行動ラベル候補付き音響信号列 16-s の各要素番号 i の要素に付与する。音響イベント判定部 112 は、この処理を入力された行動ラベル候補付き音響信号列 16-1, ..., 16-S のすべての要素 (すべての要素番号 i) について行い、その結果得られる音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, ..., 11-S を出力する。20

【0157】

音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, ..., 11-S は、音響信号列合成部 101 に入力される。以降の処理は実施例 (2)-1-1 と同じである。

【0158】

なお、音響信号列合成部 101 で音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, ..., 11-S の合成処理を行うことに代えて、音響イベント判定部 112 の前段で音響特徴量列の合成処理を行っても良い。

【0159】

< 実施例 (2)-2-1 : 図 16 >

[音響イベントラベル付き音響信号列から行動や状況を推定]

本実施例では、実施例 (2)-1-1 ~ (2)-1-3 で説明したように得られた行動 - 状況生成モデル 12、状況 - 音響イベント生成モデル 13、行動 - 音響イベント生成モデル 14 および / または行動遷移モデル 17 および / または状況遷移モデルを用い、新たに入力された音響イベントラベル付き音響信号列から行動や状況を推定する。30

【0160】

図 16 に例示するように、本実施例のモデル処理装置 1200 は、記憶部 1103 および生成モデル比較部 1201 を有する。モデル処理装置 1200 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。40

【0161】

まず生成モデル比較部 1201 に音響イベントラベル付き音響信号列 21 が入力される。生成モデル比較部 1201 は、入力された音響イベントラベル付き音響信号列 21 と、記憶部 1103 に格納された行動 - 状況生成モデル 12、状況 - 音響イベント生成モデル 13、および行動 - 音響イベント生成モデル 14 および / または行動遷移モデル 17 および / または状況遷移モデルとを比較し、音響イベントラベル付き音響信号列 21 に対し、最も適切であると判断した行動や状況、又は最も適切なものから順番に複数個の行動や状況を決定し、それらを推定結果として出力する。以下に音響イベントラベル付き音響信号列 21 と各生成モデルとの比較方法を例示する。

【0162】

10

20

30

40

50

[比較方法の例 (2) - 1]

音響イベントの分布と、行動の候補に対応する確率 P (音響イベント | 行動の候補) で表される音響イベントの分布との近さに基づいて、行動が推定されてもよい。この具体例は、実施例 (1) - 2 - 1 の [比較方法の例 (1) - 1] で説明した通りである。

【 0 1 6 3 】

また、第 2 音響イベントの分布と、状況の候補に対応する確率 P (音響イベント | 状況の候補) で表される音響イベントの分布との近さに基づいて、状況が推定されてもよい。この場合、生成モデル比較部 1201 は、式 (14) の $p(\cdot)$ と状況 - 音響イベント生成モデル 13 とを [比較方法の例 (1) - 1] と同様な方法を用いて比較することにより、入力された音響イベントラベル付き音響信号列 21 の状況を推定することもできる。例えば、式 (15) 又は (16) を用いる場合、生成モデル比較部 1201 は、 $P(\cdot)$ に $p(\cdot)$ (ただし、 $\{1, \dots, E\}$) を代入し、 $Q(\cdot)$ に式 (24) の t (ただし、 $\{1, \dots, E\}, t \in \{1, \dots, T\}$) を代入する。これにより、生成モデル比較部 1201 は、各状況 $t \in \{1, \dots, T\}$ に対応する情報量 (合計 T 個の情報量) を得る。生成モデル比較部 1201 は、各状況 $t \in \{1, \dots, T\}$ について算出された情報量のうち、最も小さな情報量に対応する状況、又は、最も小さな情報量から順番に選択した複数個の情報量に対応する複数個の状況を、音響イベントラベル付き音響信号列 21 に対応する状況として決定して出力する。なお、状況の探索範囲は予め定められたものであってもよいし、音響イベントラベル付き音響信号列 21 がこの探索範囲を特定するための情報を含み、この情報によって探索範囲が定められてもよい。探索範囲を特定するための情報の例は、状況の候補の総数 T である。このような状況の候補の総数 T は、要素番号毎に特定されてもよいし、複数の要素番号に対して共通であってもよい。この場合、探索範囲が適切に設定されていれば状況の推定精度が向上する。

【 0 1 6 4 】

生成モデル比較部 1201 は、以下のように、状況遷移モデルが表す遷移確率 $, z_i$ $-^1 z_{i-1}$ と前時刻 (要素番号 $i - 1$ の時間区間) の状況推定結果 z_{i-1} を利用して、現時刻 (要素番号 i の時間区間) の状況 $z_i = t$ を推定しても良い。

【 数 24 】

$$\arg \max_t P(z_i = t | z_{i-1}, \pi^{z_{i-1}} z_i) \quad (27-2)$$

【 0 1 6 5 】

[比較方法の例 (2) - 2]

音響イベントラベル付き音響信号列 21 が有する音響イベントの時系列の並びに対する、確率 P (音響イベント | 行動の候補) のもとでの行動の候補の尤もらしさに基づいて、行動が推定されてもよい。この具体例は、実施例 (1) - 2 - 1 の [比較方法の例 (1) - 2] で説明した通りである。

【 0 1 6 6 】

音響イベントラベル付き音響信号列 21 が有する音響イベントの時系列の並びに対する、確率 P (音響イベント | 状況の候補) のもとでの状況の候補の尤もらしさに基づいて、状況が推定されてもよい。ここでも前述のように、状況の探索範囲は予め定められたものであってもよいし、音響イベントラベル付き音響信号列 21 がこの探索範囲を特定するための情報 (例えば、状況の候補の総数) を含み、この情報によって探索範囲が定められてもよい。

【 0 1 6 7 】

以下に具体例を示す。以下の例では、入力されたラベル付き音響信号列 21 に対し、状況 - 音響イベント生成モデル 13 のもとでの状況の尤度の和や積を求める。

【 0 1 6 8 】

状況 - 音響イベント生成モデル 13 のもとでの状況の尤度の和の例

【数25】

$$P(z=t | e, \Phi, \alpha, \beta')$$

$$\begin{aligned} &= \sum_{i=1}^{N_s} P(z_i=t | e_i=\varepsilon, \Phi, \alpha, \beta') \\ &= \sum_{i=1}^{N_s} \frac{P(e_i | \Phi)}{\sum_{t=1}^T P(z_i=t, e_i=\varepsilon | \Phi, \alpha, \beta')} \end{aligned} \quad (28) \quad 10$$

【0169】

状況 - 音響イベント生成モデル13のもとでの状況の尤度の積の例

【数26】

$$P(z=t | e, \Phi, \alpha, \beta')$$

$$\begin{aligned} &= \prod_{i=1}^{N_s} P(z_i=t | e_i=\varepsilon, \Phi, \alpha, \beta') \\ &= \prod_{i=1}^{N_s} \frac{P(e_i | \Phi)}{\sum_{t=1}^T P(z_i=t, e_i=\varepsilon | \Phi, \alpha, \beta')} \end{aligned} \quad (29) \quad 20$$

【0170】

ただし、式(28)(29)の e_i は、入力された音響イベントラベル付き音響信号列21の要素番号*i*に対応する音響イベントラベルが表す音響イベントを表す。式(28)(29)は、式(24)(25)(27)の確率 $P(e'_s | t, t_a, \dots)$ と、入力された音響イベントラベル付き音響信号列21の e_i とから算出できる。

【0171】

状況の探索範囲が予め定められたものである場合には、その探索範囲に属する各状況の候補*t*について上記の尤度が算出され、状況の探索範囲が音響イベントラベル付き音響信号列21に含まれる情報（例えば、状況の候補の総数等）によって特定される場合には、その探索範囲に属する各状況の候補*t*について上記の尤度が算出される。生成モデル比較部1201は、各状況の候補について算出した尤度のうち、最も尤度の高い状況の候補、又は、最も尤度の高いものから順番に選択した複数個の状況の候補を、入力された音響イベントラベル付き音響信号列21に対応する状況として決定して出力する。

【0172】

生成モデル比較部1201は、以下のように、行動 - 音響イベント生成モデル14のもとでの状況の尤度の和や積および、状況遷移モデルが表す遷移確率 $P(z_{i-1} | z_i)$ と前時刻（要素番号*i*-1の時間区間）の状況推定結果 z_{i-1} を利用して、現時刻（要素番号*i*の時間区間）の状況 $z_i = t$ を推定しても良い

【0173】

行動 - 音響イベント生成モデル14のもとでの状況の尤度の和および、状況遷移モデルと前時刻の状況推定結果を利用する場合の例

【数27】

$$\begin{aligned}
 & P(z=t | e, \Phi, \alpha, \beta', z_{i-1}, \pi'^{z_{i-1}} z_i) \\
 &= \sum_{i=1}^{N_s} P(z_i = t | e_i = \varepsilon, \Phi, \alpha, \beta', z_{i-1}, \pi'^{z_{i-1}} z_i) \\
 &= \sum_{i=1}^{N_s} \frac{P(e_i | \Phi)}{\sum_{t=1}^T P(z_i = t, e_i = \varepsilon | \Phi, \alpha, \beta')} \\
 &\quad \times P(z_i = t | z_{i-1}, \pi'^{z_{i-1}} z_i) \quad (28-2)
 \end{aligned}$$

10

【0174】

行動 - 音響イベント生成モデル14のもとでの状況の尤度の積および、状況遷移モデルと前時刻の状況推定結果を利用する場合の例

【数28】

$$\begin{aligned}
 & P(z=t | e, \Phi, \alpha, \beta', z_{i-1}, \pi'^{z_{i-1}} z_i) \\
 &= \prod_{i=1}^{N_s} P(z_i = t | e_i = \varepsilon, \Phi, \alpha, \beta', z_{i-1}, \pi'^{z_{i-1}} z_i) \\
 &= \prod_{i=1}^{N_s} \frac{P(e_i | \Phi)}{\sum_{t=1}^T P(z_i = t, e_i = \varepsilon | \Phi, \alpha, \beta')} \\
 &\quad \times P(z_i = t | z_{i-1}, \pi'^{z_{i-1}} z_i) \quad (29-2)
 \end{aligned}$$

20

【0175】

30

なお、上述の行動を決定する処理と状況を決定する処理とが、互いに異なる処理部で実行されてもよい。

【0176】

<実施例(2)-2-2:図17>

[音響信号列から行動および状況を推定]

本実施例では、実施例(2)-1-1～(2)-1-3で説明したように得られた行動 - 状況生成モデル12、状況 - 音響イベント生成モデル13、行動 - 音響イベント生成モデル14および/または行動遷移モデル17および/または状況遷移モデルを用い、新たに入力された音響信号列から行動や状況を推定する。

【0177】

40

図17に例示するように、本実施例のモデル処理装置1210は、特徴量算出部211、音響イベント判定部212、音響イベントモデルDB113、記憶部1103、生成モデル比較部1201を有する。モデル処理装置1210は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0178】

まず特徴量算出部211に音響信号列22が入力される。特徴量算出部211は、音響信号列22から音響特徴量列(ベクトル)を算出して出力する。例えば特徴量算出部211は、前述した特徴量算出部111と同じ方法で音響特徴量列を算出する。

【0179】

音響イベント判定部212は、前述した音響イベント判定部112と同じ方法で、特徴

50

量算出部 211 から出力された音響特徴量列と、音響イベントモデル DB113 に記憶されている複数の音響イベントモデルとをそれぞれ比較し、音響信号列 22 の全ての要素について音響イベントを決定する。音響イベント判定部 212 は、決定した音響イベントを表す音響イベントラベルを音響信号列 22 の各要素に付することで、音響イベントラベル付き音響信号列 21 を生成して出力する。

【0180】

音響イベントラベル付き音響信号列 21 は、生成モデル比較部 1201 に入力される。以降の処理は実施例(2)-2-1と同じである。

【0181】

< 実施例(2)-2-3 : 図 18 >

10

[音響特徴量から行動や状況を推定]

本実施例では、実施例(2)-1-1～(2)-1-3で説明したように得られた行動 - 状況生成モデル 12、状況 - 音響イベント生成モデル 13、行動 - 音響イベント生成モデル 14、および/または行動遷移モデル 17 および/または状況遷移モデルを用い、新たに入力された音響特徴量列から行動や状況を推定する。

【0182】

図 18 に例示するように、本実施例のモデル処理装置 1220 は、音響イベント判定部 212、音響イベントモデル DB113、記憶部 1103、生成モデル比較部 1201 を有する。モデル処理装置 1220 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

20

【0183】

まず音響イベント判定部 212 に音響特徴量列 23 が入力される。音響イベント判定部 212 は、前述した音響イベント判定部 112 と同じ方法で、音響特徴量列 23 の音響特徴量列と、音響イベントモデル DB113 に記憶されている複数の音響イベントモデルとをそれぞれ比較し、音響特徴量列 23 が含む音響信号列の全ての要素について音響イベントを決定する。音響イベント判定部 212 は、決定した各要素の音響イベントを表す音響イベントラベルを、音響特徴量列 23 が含む音響信号列に付することで、音響イベントラベル付き音響信号列 21 を生成して出力する。

【0184】

音響イベントラベル付き音響信号列 21 は、生成モデル比較部 1201 に入力される。以降の処理は実施例(2)-2-1と同じである。

30

【0185】

< 実施例(2)-3-1 : 図 19 >

[音響イベントラベル付き音響信号列から行動や状況を推定すると共に、音響イベントラベルおよび行動ラベル候補付き音響信号列から行動 - 音響イベント生成モデル、状況 - 音響イベント生成モデル、行動 - 状況生成モデル、および行動遷移モデルを算出]

本実施例は実施例(2)-1-1と実施例(2)-2-1の組み合わせである。本実施例では、音響イベントラベル付き音響信号列 21 を入力として行動や状況を推定することに加え、音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, ..., 1-1-S を入力とし、行動 - 状況生成モデル 12、状況 - 音響イベント生成モデル 13、行動 - 音響イベント生成モデル 14 の少なくとも何れかの算出も行う。さらに行動遷移モデル 17 や行動ラベル 18 や状況ラベル 19 の生成を行ってもよい。さらに、推定された行動を表す行動ラベル候補を、音響イベントラベル付き音響信号列 21 が含む音響信号列の各要素に付与することで、音響イベントラベルおよび行動ラベル候補付き音響信号列を得、それを生成モデルの更新に利用することもできる。

40

【0186】

図 19 に例示するように、本実施例のモデル処理装置 1300 は、記憶部 1103, 1303、音響信号列合成部 101、モデル化部 1102、および生成モデル比較部 1201 を有する。モデル処理装置 1300 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

50

【0187】

記憶部303には、前述した音響イベントラベルおよび行動ラベル候補付き音響信号列11-1, . . . , 11-S、および前述した音響イベントラベル付き音響信号列21が格納されている。

【0188】

音響イベントラベルおよび行動ラベル候補付き音響信号列11-1, . . . , 11-Sは音響信号列合成部101に入力される。音響信号列合成部101およびモデル化部1102は、実施例(2)-1-1で説明したように、音響イベントラベルおよび行動ラベル候補付き音響信号列11-1, . . . , 11-Sから、行動・状況生成モデル12、状況-音響イベント生成モデル13、および行動-音響イベント生成モデル14の少なくとも何れかを生成し、それらを記憶部1103に格納する。また、実施例(2)-1-1で説明したように、モデル化部1102が行動遷移モデル17を生成して記憶部1103に格納してもよいし、行動ラベル18を生成して記憶部103に格納してもよいし、状況ラベル19を生成して記憶部1103に格納してもよい。

10

【0189】

音響イベントラベル付き音響信号列21は生成モデル比較部1201に入力される。音響イベントラベル付き音響信号列21は、実施例(2)-2-1で説明したように、入力された音響イベントラベル付き音響信号列21と、記憶部1103に格納された行動・状況生成モデル12、状況-音響イベント生成モデル13、行動-音響イベント生成モデル14、および/または行動遷移モデル17および/または状況遷移モデルとを比較し、音響イベントラベル付き音響信号列21に対し、最も適切であると判断した行動や状況、又は最も適切なものから順番に複数個の行動や状況を選択し、それらを推定結果として出力する。

20

【0190】

さらに生成モデル比較部1201は、決定した行動を表す行動ラベル候補を、音響イベントラベル付き音響信号列21の対応する各要素に付与し、それによって新たな音響イベントラベルおよび行動ラベル候補付き音響信号列21'を生成して出力してもよい。この場合、音響イベントラベルおよび行動ラベル候補付き音響信号列21'が、音響イベントラベルおよび行動ラベル候補付き音響信号列11-1, . . . , 11-Sとともに音響信号列合成部101に入力されてもよい。この場合、音響信号列合成部101は、入力されたすべての音響イベントラベルおよび行動ラベル候補付き音響信号列11-1, . . . , 11-S, 21'を時系列方向につなぎ合わせ(合成)、それによって一つの音響イベントラベルおよび行動ラベル候補付き音響信号列11'('第3音響イベントラベル列'および'第3行動ラベル候補')を得て出力する。モデル化部1102は、実施例(2)-1-1のラベル付き音響信号列11('音響イベントラベル列'および'行動ラベル候補')に代えて、音響イベントラベルおよび行動ラベル候補付き音響信号列11'('第3音響イベントラベル列'および'第3行動ラベル候補')を用い、再度、実施例(2)-1-1で説明したように、行動・状況生成モデル12、状況-音響イベント生成モデル13、および行動-音響イベント生成モデル14の少なくとも何れかを生成し、それを記憶部1103に格納する。さらにモデル化部1102が、ラベル付き音響信号列11に代えて、入力された音響イベントラベルおよび行動ラベル候補付き音響信号列11'を用い、再度、実施例(2)-1-1で説明したように、行動遷移モデル17を生成して記憶部1103に格納してもよいし、行動ラベル18や状況ラベル19を生成して記憶部1103に格納してもよい。

30

40

【0191】

また、生成モデル比較部1201の処理およびモデル化部1102の処理のどちらを行っても良い。ただし、モデル化部1102の処理を行う前に生成モデル比較部1201の処理を行う場合、記憶部1103に予め得られた各生成モデルおよび/または行動遷移モデル17および/または状況遷移モデルが格納されている必要がある。

【0192】

50

また、音響イベントラベルおよび行動ラベル候補付き音響信号列 21' が、新たに入力された音響イベントラベルおよび行動ラベル候補付き音響信号列とともに音響信号列合成部 101 に入力されてもよい。音響信号列合成部 101 は、これらを時系列方向につなぎ合わせ（合成）、それによって一つの音響イベントラベルおよび行動ラベル候補付き音響信号列（「第 3 音響イベントラベル列」および「第 3 行動ラベル候補」）を得て、モデル化部 1102 に送出してもよい。

その他の処理は実施例(2)-1-1 および実施例(2)-2-1 と同様とする。

【0193】

<実施例(2)-3-2：図 20 >

[音響信号列から行動、状況を推定すると共に、行動ラベル候補付き音響信号列から行動 - 音響イベント生成モデル、状況 - 音響イベント生成モデル、行動 - 状況生成モデル、および行動遷移モデルを算出]

本実施例は、実施例(2)-1-2 と実施例(2)-2-2 の組み合わせである。

本実施例では、行動ラベル候補付き音響信号列 15-1, ..., 15-S を入力として、学習によって、行動 - 状況生成モデル 12、状況 - 音響イベント生成モデル 13、行動 - 音響イベント生成モデル 14 の少なくとも何れかを算出する。さらに行動遷移モデル 17 や行動ラベル 18 や状況ラベル 19 の生成を行ってもよい。さらに本実施例では、行動 - 状況生成モデル 12、状況 - 音響イベント生成モデル 13、行動 - 音響イベント生成モデル 14、および / または行動遷移モデル 17 および / または状況遷移モデルを用い、新たに入力された音響信号列 22 から行動や状況を推定する。

【0194】

図 20 に例示するように、本実施例のモデル処理装置 1310 は、特微量算出部 111-1, ..., 111-S, 211、音響イベント判定部 112-1, ..., 112-S, 212、音響イベントモデル DB113、および前述のモデル処理装置 1300（図 19 参照）を有する。

【0195】

行動ラベル候補付き音響信号列 15-1, ..., 15-S は、それぞれ特微量算出部 111-1, ..., 111-S に入力される。特微量算出部 111-1, ..., 111-S は、実施例(1)-1-2 で説明したように、行動ラベル候補付き音響信号列 15-1, ..., 15-S から、それぞれ音響特微量列を得て出力する。音響イベント判定部 112-1, ..., 112-S は、それぞれ、前述の音響イベント判定部 112 と同様に、入力された音響特微量列と音響イベントモデル DB113 に記憶されている複数の音響イベントモデルとから、音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, ..., 11-S を生成して出力する。

【0196】

音響信号列 22 は特微量算出部 211 に入力される。特微量算出部 211 は、実施例(1)-2-2 で説明したように、音響信号列 22 から音響特微量列（ベクトル）を算出して出力する。音響イベント判定部 212 は、実施例(1)-2-2 で説明したように、入力された音響特微量列と音響イベントモデル DB113 に記憶されている複数の音響イベントモデルとから、音響イベントラベル付き音響信号列 21 を生成して出力する。

【0197】

音響イベントラベルおよび行動ラベル候補付き音響信号列 11-1, ..., 11-S、音響イベントラベル付き音響信号列 21 は、記憶部 1303（図 19）に格納される。以降の処理は実施例(2)-3-1 と同じである。

【0198】

<実施例(2)-3-3：図 21 >

[音響特微量から行動や状況を推定すると共に、行動ラベル候補付き音響信号列から行動 - 音響イベント生成モデル、状況 - 音響イベント生成モデル、行動 - 状況生成モデル、および行動遷移モデルを算出]

本実施例は、実施例(2)-1-3 と実施例(2)-2-3 の組み合わせである。

10

20

30

40

50

本実施例では、行動ラベル候補付き音響特徴量列 16 - 1, . . . , 16 - S を入力として、学習によって、行動 - 状況生成モデル 12、状況 - 音響イベント生成モデル 13、行動 - 音響イベント生成モデル 14 の少なくとも何れかを算出する。行動遷移モデル 17 や行動ラベル 18 や状況ラベル 19 の生成を行ってもよい。さらに本実施例では、行動 - 状況生成モデル 12、状況 - 音響イベント生成モデル 13、行動 - 音響イベント生成モデル 14 を用い、新たに入力された音響特徴量列 23 から行動や状況を推定する。

【0199】

図 21 に例示するように、本実施例のモデル処理装置 1320 は、音響イベント判定部 112 - 1, . . . , 112 - S, 212、音響イベントモデル DB113、および前述のモデル処理装置 1300 (図 19 参照) を有する。

10

【0200】

行動ラベル候補付き音響特徴量列 16 - 1, . . . , 16 - S は、それぞれ音響イベント判定部 112 - 1, . . . , 112 - S に入力される。音響イベント判定部 112 - 1, . . . , 112 - S は、それぞれ、前述の音響イベント判定部 112 と同様に、入力された音響特徴量列と音響イベントモデル DB113 に記憶されている複数の音響イベントモデルとから、音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, . . . , 11 - S を生成して出力する。

【0201】

音響特徴量列 23 は音響イベント判定部 212 に入力される。音響イベント判定部 212 は、実施例 (1) - 2 - 3 と同様に、入力された音響特徴量列 23 の音響特徴量列と音響イベントモデル DB113 に記憶されている複数の音響イベントモデルとから、音響イベントラベル付き音響信号列 21 を生成して出力する。

20

【0202】

音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, . . . , 11 - S、音響イベントラベル付き音響信号列 21 は、記憶部 303 (図 19) に格納される。以降の処理は実施例 (2) - 3 - 1 と同じである。

【0203】

< 実施例 (2) - 4 - 1 : 図 22 >

[音響イベントラベル付き音響信号列から行動、状況を推定すると共に、音響イベントラベルおよび行動ラベル候補付き音響信号列から行動 - 音響イベント生成モデル状況 - 音響イベント生成モデル、行動 - 状況生成モデル、および行動遷移モデルを算出。特別なシンボルを用いる]

30

本実施例は実施例 (2) - 3 - 1 の変形である。

本実施例では、音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, . . . , 11 - S、および音響イベントラベル付き音響信号列 21 を入力とし、行動 - 状況生成モデル 12、状況 - 音響イベント生成モデル 13、行動 - 音響イベント生成モデル 14 の少なくとも何れかを算出する。また行動遷移モデル 17 の生成や行動や状況の推定を行ってもよい。さらに、推定された行動を表す行動ラベル候補を、音響イベントラベル付き音響信号列 21 が含む音響信号列の各要素に付与することで、音響イベントラベルおよび行動ラベル候補付き音響信号列を得、それを生成モデルの更新に利用することもできる。

40

【0204】

図 22 に例示するように、本実施例のモデル処理装置 1400 は、記憶部 1103、1303、音響信号列合成部 401、モデル化部 1402、および生成モデル比較部 1403 を有する。モデル処理装置 1400 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0205】

音響信号列合成部 401 に、音響イベントラベルおよび行動ラベル候補付き音響信号列 11 - 1, . . . , 11 - S および音響イベントラベル付き音響信号列 21 が入力される。音響信号列合成部 401 は、音響イベントラベルおよび行動ラベル候補付き音響信号列

50

1 1 - 1 , . . . , 1 1 - S 、音響イベントラベル付き音響信号列 2 1 を時系列方向につなぎ合わせ、それによって 1 つの音響イベントラベルおよび行動ラベル候補付き音響信号列 4 1 (以下、単に「ラベル付き音響信号列 4 1」という)を得て出力する。実施例(1) - 4 - 1 で説明したように、ラベル付き音響信号列 4 1 が含む音響イベントの時系列の並びは、行動ラベル候補によって行動の候補が表された第 4 音響イベントの時系列の並び、および行動ラベル候補によって行動の候補が表されていない第 5 音響イベントの時系列の並びを含む。ラベル付き音響信号列 4 1 は、モデル化部 1 4 0 2 に入力される。なお、予め、音響イベントラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S 、音響イベントラベル付き音響信号列 2 1 からラベル付き音響信号列 4 1 が得られている場合には、音響信号列合成部 4 0 1 を経由することなく、ラベル付き音響信号列 4 1 がそのままモデル化部 1 4 0 2 に入力されてもよい。10

【 0 2 0 6 】

モデル化部 1 4 0 2 は、実施例(2) - 1 - 1 のラベル付き音響信号列 1 1 に代えて、ラベル付き音響信号列 4 1 を用い、実施例(2) - 1 - 1 のモデル化部 1 1 0 2 と同じ方法で、行動 - 状況生成モデル 1 2 、状況 - 音響イベント生成モデル 1 3 、および行動 - 音響イベント生成モデル 1 4 の少なくとも何れかを算出する。行動遷移モデル 1 7 や行動ラベル 1 8 の生成を行ってもよい。ただし、モデル化部 1 4 0 2 は、実施例(1) - 4 - 1 のモデル化部 4 0 2 と同様、上述の「第 5 音響イベント」を生じさせた行動の候補として特別なシンボルを用いて学習処理を行う。20

【 0 2 0 7 】

すなわちモデル化部 1 4 0 2 は、状況の候補が音響イベントを生成する確率 P (音響イベント | 状況の候補) 、行動の候補が状況の候補を生成する確率 P (状況の候補 | 行動の候補) 、およびこの特別なシンボルが状況の候補を生成する確率 P (状況の候補 | 特別なシンボル) を得る。またモデル化部 1 4 0 2 は、確率 P (音響イベント | 状況の候補) と確率 P (状況の候補 | 行動の候補) とから、行動の候補が音響イベントを生成する確率 P (音響イベント | 行動の候補) を得、確率 P (音響イベント | 状況の候補) と確率 P (状況の候補 | 特別なシンボル) とから、特別なシンボルが音響イベントを生成する確率 P (音響イベント | 特別なシンボル) を得る。行動 - 状況生成モデル 1 2 は、確率 P (状況の候補 | 行動の候補) , P (状況の候補 | 特別なシンボル) に対応し、状況 - 音響イベント生成モデル 1 3 は、確率 P (音響イベント | 状況の候補) に対応し、行動 - 音響イベント生成モデル 1 4 は、確率 P (音響イベント | 行動の候補) , P (音響イベント | 特別なシンボル) に対応する。モデル化部 1 4 0 2 で算出された、行動 - 状況生成モデル 1 2 、状況 - 音響イベント生成モデル 1 3 、および行動 - 音響イベント生成モデル 1 4 は、記憶部 1 1 0 3 に格納される。行動遷移モデル 1 7 や行動ラベル 1 8 や状況ラベル 1 9 の生成を行った場合には、それらも記憶部 1 0 3 に格納される。30

【 0 2 0 8 】

生成モデル比較部 1 4 0 3 には、行動 - 音響イベント生成モデル 1 4 および/または行動遷移モデル 1 7 が入力される。生成モデル比較部 1 4 0 3 は、入力された行動 - 音響イベント生成モデル 1 4 に対応する、確率 P (音響イベント | 行動) と確率 P (音響イベント | 特別なシンボル) とを比較し、特別なシンボルに対応する行動の内容を推定し、それらの推定結果を出力する。また、生成モデル比較部 1 4 0 3 は、入力された行動遷移モデル 1 7 に対応する、遷移確率 P (現時刻の行動の候補 | 前時刻の行動の候補) と遷移確率 P (特別なシンボル | 前時刻の行動の候補) とを比較し、特別なシンボルに対応する行動の内容を推定し、それらの推定結果を出力してもよい。例えば、生成モデル比較部 1 4 0 3 は、音響イベントラベル付き音響信号列 2 1 に対し、最も適切であると判断した行動、又は最も適切なものから順番に複数個の行動を選択し、それらを推定結果として出力する。40

【 0 2 0 9 】

[比較方法の例示]

生成モデル比較部 1 4 0 3 による比較は、例えば、実施例(1) - 2 - 1 で説明した情50

報量基準に基づいて行うことができる。例えば、式(15)又は(16)の情報量基準を用い、確率P(音響イベント|行動の候補)と確率P(音響イベント|特別なシンボル)とを比較する場合、生成モデル比較部1403は、P()に特別なシンボル $a_i = a' \{ 1, \dots, A \}$ に対応する式(27)の確率(ただし、 $i = \{ 1, \dots, E \}$)を代入し、Q()に他の行動 $a_i = a'' \neq a'$ (ただし、 $a'' \in \{ 1, \dots, A \}$)に対応する式(27)の確率(ただし、 $i = \{ 1, \dots, E \}$)を代入する。これにより、生成モデル比較部1403は、他の行動 $a'' \neq a'$ (ただし、 $a'' \in \{ 1, \dots, A \}$)に対応する情報量(合計A-1個の情報量)を得る。

【0210】

生成モデル比較部1403は、各行動 a'' について算出された情報量のうち、最も小さな情報量に対応する行動の候補、又は、最も小さな情報量から順番に選択した複数個の情報量に対応する複数個の行動の候補を、音響イベントラベル付き音響信号列21に対応する行動として決定して出力する。 10

【0211】

生成モデル比較部1403は、以上のように決定した行動を表す行動ラベル候補を、音響イベントラベル付き音響信号列21の対応する各要素に付与し、それによって新たな音響イベントラベルおよび行動ラベル候補付き音響信号列41'を生成して出力してもよい。この場合、音響イベントラベルおよび行動ラベル候補付き音響信号列41'が、音響イベントラベルおよび行動ラベル候補付き音響信号列11-1, ..., 11-Sとともに音響信号列合成部401に入力されてもよい。この場合、音響信号列合成部401は、入力されたすべての音響イベントラベルおよび行動ラベル候補付き音響信号列11-1, ..., 11-S, 41'を時系列方向につなぎ合わせ(合成)、それによって一つの音響イベントラベルおよび行動ラベル候補付き音響信号列41''を得て出力する。モデル化部1402は、実施例(2)-1-1のラベル付き音響信号列11に代えて、音響イベントラベルおよび行動ラベル候補付き音響信号列41''を用い、再度、実施例(2)-1-1で説明したように、行動-状況生成モデル12、状況-音響イベント生成モデル13、および行動-音響イベント生成モデル14の少なくとも何れかを生成し、それを記憶部1103に格納する。さらにモデル化部1402が、入力された音響イベントラベルおよび行動ラベル候補付き音響信号列41''を用い、再度、実施例(2)-1-1で説明したように、行動遷移モデル17を生成して記憶部1103に格納してもよいし、行動ラベル18や状況ラベル19を生成して記憶部1103に格納してもよい。 20

【0212】

また、音響イベントラベルおよび行動ラベル候補付き音響信号列21'が、新たに入力された音響イベントラベルおよび行動ラベル候補付き音響信号列とともに音響信号列合成部401に入力されてもよい。音響信号列合成部401は、これらを時系列方向につなぎ合わせ(合成)、それによって一つの音響イベントラベルおよび行動ラベル候補付き音響信号列を得て、モデル化部1402に送出し、モデル化部102が上述の処理を行ってもよい。

その他の処理は実施例(2)-1-1、実施例(2)-2-1、および実施例(2)-3-1と同じである。 40

【0213】

<実施例(2)-4-2：図20>

[音響信号列から行動、状況を推定すると共に、行動ラベル候補付き音響信号列から行動-音響イベント生成モデル、状況-音響イベント生成モデル、行動-状況生成モデル、および行動遷移モデルを算出。特別なシンボルを用いる]

本実施例は実施例(2)-3-2の変形である。

本実施例では、行動ラベル候補付き音響信号列15-1, ..., 15-Sおよび音響信号列22を入力として、学習によって、行動-状況生成モデル12、状況-音響イベント生成モデル13、行動-音響イベント生成モデル14の少なくとも何れかを算出する。さらに行動遷移モデル17や行動ラベル18の生成を行ってもよい。さらに本実施例では 50

、行動 - 状況生成モデル 1 2 、状況 - 音響イベント生成モデル 1 3 、行動 - 音響イベント生成モデル 1 4 を用い、音響信号列 2 2 に対応する行動や状況を推定する。

【 0 2 1 4 】

図 2 0 に例示するように、本実施例のモデル処理装置 1 4 1 0 は、特徴量算出部 1 1 1 - 1 , . . . , 1 1 1 - S , 2 1 1 、音響イベント判定部 1 1 2 - 1 , . . . , 1 1 2 - S , 2 1 2 、音響イベントモデル DB 1 1 3 、および前述のモデル処理装置 1 4 0 0 (図 2 2 参照) を有する。

【 0 2 1 5 】

行動ラベル候補付き音響信号列 1 5 - 1 , . . . , 1 5 - S は、それぞれ特徴量算出部 1 1 1 - 1 , . . . , 1 1 1 - S に入力され、音響信号列 2 2 は特徴量算出部 2 1 1 に入力される。特徴量算出部 1 1 1 - 1 , . . . , 1 1 1 - S , 2 1 1 、音響イベント判定部 1 1 2 - 1 , . . . , 1 1 2 - S , 2 1 2 は、実施例 (2) - 3 - 2 で説明したように、音響イベントラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S 、音響イベントラベル付き音響信号列 2 1 を生成して出力する。音響イベントラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S 、音響イベントラベル付き音響信号列 2 1 は、記憶部 1 3 0 3 (図 2 2) に格納される。以降の処理は実施例 (2) - 4 - 1 と同じである。

【 0 2 1 6 】

< 実施例 (2) - 4 - 3 : 図 2 1 >

[音響特徴量から行動、状況を推定すると共に、行動ラベル候補付き音響信号列から行動 - 音響イベント生成モデル、状況 - 音響イベント生成モデル、行動 - 状況生成モデル、および行動遷移モデルを算出。特別なシンボルを用いる]

本実施例は実施例 (2) - 3 - 3 の変形である。

本実施例では、行動ラベル候補付き音響特徴量列 1 6 - 1 , . . . , 1 6 - S および音響特徴量列 2 3 を入力として、学習によって、行動 - 状況生成モデル 1 2 、状況 - 音響イベント生成モデル 1 3 、行動 - 音響イベント生成モデル 1 4 の少なくとも何れかを算出する。行動遷移モデル 1 7 や行動ラベル 1 8 や状況ラベル 1 9 の生成を行ってもよい。さらに本実施例では、行動 - 状況生成モデル 1 2 、状況 - 音響イベント生成モデル 1 3 、行動 - 音響イベント生成モデル 1 4 を用い、音響特徴量列 2 3 に対応する行動や状況を推定する。

【 0 2 1 7 】

図 2 1 に例示するように、本実施例のモデル処理装置 1 4 2 0 は、音響イベント判定部 1 1 2 - 1 , . . . , 1 1 2 - S , 2 1 2 、音響イベントモデル DB 1 1 3 、および前述のモデル処理装置 1 4 0 0 (図 2 2 参照) を有する。

【 0 2 1 8 】

行動ラベル候補付き音響特徴量列 1 6 - 1 , . . . , 1 6 - S は、それぞれ音響イベント判定部 1 1 2 - 1 , . . . , 1 1 2 - S に入力され、音響特徴量列 2 3 は、音響イベント判定部 2 1 2 に入力される。音響イベント判定部 1 1 2 - 1 , . . . , 1 1 2 - S , 2 1 2 は、実施例 (2) - 3 - 3 で説明したように、音響イベントラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S 、音響イベントラベル付き音響信号列 2 1 を生成して出力する。音響イベントラベルおよび行動ラベル候補付き音響信号列 1 1 - 1 , . . . , 1 1 - S 、音響イベントラベル付き音響信号列 2 1 は、記憶部 1 3 0 3 (図 2 2) に格納される。以降の処理は実施例 (2) - 4 - 1 と同じである。

【 0 2 1 9 】

< 実施例 (3) - 1 - 1 : 図 2 3 >

[音響イベントラベル付き音響信号から音響信号 - 状況生成モデルおよび状況 - 音響イベントの生成モデルを、状況の時間変化を考慮に入れて算出し、同時に、状況を分析]

本実施例では、状況の時間変化を考慮に入れ、音響イベントラベル付き音響信号から、状況 - 音響イベント生成モデルを算出する。同時に、音響信号 - 状況生成モデルを算出してもよく、状況の時間遷移を表す状況遷移モデルを算出しててもよく、状況の分析も行って

10

20

30

40

50

もよい。

【0220】

すなわち、本実施例では、各時間区間の音響信号を生じさせた音響イベントの時系列の並びを表す音響イベントラベル列、および音響イベントを生じさせた状況の候補の総数を用い、「音響イベントの時系列の並びにおける、状況の候補に対応する音響イベントの組み合わせと状況の候補の時系列の並びとの同時事後確率」を最大化するための処理を含む学習処理を行う。例えば、この学習処理は、状況の時間遷移に基づく状況の出現確率および、状況における音響イベントの出現確率に基づいて、音響イベントの時系列の並びにおける、状況の候補に対する音響イベントの組み合わせと状況の候補の時系列の並びとの同時事後確率を最大化するための処理である。例えば、音響イベントの時系列の並びからなる音響イベントの組み合わせに対する尤度関数 P （音響イベントの組み合わせ | 状況の候補の時系列の並び）を最大にする処理が行われる。それにより、状況の候補が音響イベントを生成する確率 P （音響イベント | 状況の候補）を「状況 - 音響イベント生成モデル」として得る。また、この学習処理によって、音響信号列が状況を生成する確率 P （状況 | 音響信号列）を「音響信号 - 状況生成モデル」として得てもよいし、状況の候補の時系列の遷移確率を「状況遷移モデル」として得てもよい。また、この学習処理の過程で何れかの状況の候補を選択し、各音響イベントがどの状況によって生成されたのかを分析してもよい。この場合には選択した状況の候補を表す状況ラベルを出力してもよい。

10

【0221】

図23に例示するように、本実施例のモデル処理装置2100は、音響信号列合成部2101、モデル化部2102、および記憶部2103を有する。モデル処理装置2100は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

20

【0222】

まず音響信号列合成部2101に、音響イベントラベル付き音響信号列2011-1, ..., 2011-S（ただし、Sは1以上の整数）が入力される。各音響イベントラベル付き音響信号列2011-s（ただし、s ∈ {1, ..., S}）は、時系列の音響信号列2011a-s、各音響信号列2011a-sに対応する音響信号列番号、前述の時間区間にごとに区分された音響信号列の各要素に対応する要素番号、および当該時間区間にごとに決定されて付与された音響イベントラベルを含む。各音響信号列2011a-sは音を表すデジタル信号列である。音響イベントラベルは、音響信号列の各要素に対応する音響イベントを表すラベルであり、音響信号列の要素ごとに付与される。1個の音響信号列番号には、1個以上の要素番号が対応する。

30

【0223】

複数個の音響イベントラベル付き音響信号列2011-1, ..., 2011-S（以下、単に「ラベル付き音響信号列2011-1, ..., 2011-S」という）が音響信号列合成部2101に入力された場合、音響信号列合成部2101は、それらを時系列方向につなぎ合わせ、それによって1つの音響イベントラベル付き音響信号列2011（以下、単に「ラベル付き音響信号列2011」という）を得て出力する（合成処理）。音響信号列合成部2101に1つの音響信号列2011-1のみが入力された場合、音響信号列合成部2101はそれをラベル付き音響信号列2011として出力する。音響信号列合成部2101から出力された音響イベントラベル付き音響信号列は、モデル化部2102に入力される。なお、音響信号列合成部2101を経由することなく、1つラベル付き音響信号列2011がそのままモデル化部2102に入力されてもよい。また、ラベル付き音響信号列2011が含む合成された音響信号列2011a-1, ..., 2011a-Sの並びを音響信号列2011aと表記する。

40

【0224】

モデル化部2102は、以下の手順に従って、入力されたラベル付き音響信号列2011から、音響信号 - 状況生成モデル12、および状況 - 音響イベント生成モデル13の少なくとも何れかを生成し、それらを記憶部2103に格納する。モデル化部102は、状

50

況の候補の遷移を表す状況遷移モデル 2017 を生成してもよいし、状況ラベル 19 を生成してもよい。生成された状況遷移モデル 2017 や状況ラベル 19 も記憶部 2103 に格納される。

【0225】

[状況から音響イベントが生成される過程の論理的説明]

モデル化部 2102 は、ラベル付き音響信号列 2011 から、音響信号列が状況の候補を生成する確率 $P(\text{状況の候補} | \text{音響信号列})$ としたときの音響信号 - 状況生成モデル 12 と、状況の候補が音響イベントを生成する確率 $P(\text{音響イベント} | \text{状況の候補})$ としたときの状況 - 音響イベント生成モデル 13 の少なくとも何れかを算出する。つまり本実施例では、音響信号列が潜在的な場の状況の生成確率を規定し、状況が音響イベントの生成確率を規定すると考え、これらの関係を各生成モデルとして記述する。また、状況の候補の時系列の遷移確率を算出し、状況遷移モデル 2017 を得てもよい。

【0226】

合成された音響信号列における状況の候補の生成確率 $'$ 、状況の候補における音響イベントの生成確率 $'$ 、音響信号列の集合 $'$ が与えられた場合における、音響イベントの列 e の生成確率 $P(e | ' , ' , ' , ' , ')$ は以下の通りである。

【数 29】

$$P(e | \Pi', \Theta', \Phi', \Omega') = \prod_{s=1}^S P(e'_s | \Pi', \Theta', \Phi') \quad (31)$$

10

20

ただし、 S は合成された音響信号列 2011a に含まれる音響信号列 2011a - s (ただし、 $s \in \{1, \dots, S\}$) の個数、 T は状況の候補の総数、 E は音響イベントの総数 (音響イベントの種類の個数)、 e はラベル付き音響信号列 2011 に与えられた音響イベントの列 (ベクトル)、 $'$ は各時間区間の状況の候補 t (ただし、 $t \in \{1, \dots, T\}$) が次の時間区間で状況の候補 t' (ただし、 $t' \in \{1, \dots, T\}$) に遷移する確率を (t, t') 要素とする $T \times T$ 行列、 $'$ は音響信号列 2011a - s (ただし、 $s \in \{1, \dots, S\}$) が状況の候補 t (ただし、 $t \in \{1, \dots, T\}$) を生成する確率 $P(t | s)$ を (s, t) 要素とする $S \times T$ 行列、 $'$ は状況の候補 t (ただし、 $t \in \{1, \dots, T\}$) が音響イベント (ただし、 $e \in \{1, \dots, E\}$) を生成する確率 $P(e | t)$ を (t, e) 要素とする $T \times E$ 行列、 e'_s は音響信号列 2011a - s に与えられた音響イベントの列 (N_s 次元ベクトル: N_s は音響信号列 2011a - s に対応する音響イベントの個数)、 $'$ は音響信号列 2011a と成り得る列の集合を表す。

【0227】

ラベル付き音響信号列 2011 に対応する状況の候補 t の遷移確率 $_t$ の事前分布のパラメータ (Dirichlet 分布に従うものとする) $'$ 、各音響信号列 2011a - s における状況の候補 t の生成確率 $_s$ の事前分布のパラメータ (Dirichlet 分布に従うものとする) $'$ 、状況の候補 t における音響イベント e の生成確率 $_t$ の事前分布のパラメータ (Dirichlet 分布に従うものとする) $'$ が与えられたときの、音響イベントの列 e'_s の生成確率 $P(e'_s, ' , ' , ' , ' | ' , ' , ')$ は、以下の通りである。

【数 30】

$$\begin{aligned} & P(e'_s, \Pi', \Theta', \Phi' | \gamma', \alpha', \beta') \\ &= P(\pi_t | \gamma') P(\theta_s | \alpha') P(\phi_t | \beta') \\ & \times \prod_{i=1}^{N_s} P(e_i | z_i, \phi_t) P(z_i | \theta_s, \pi_{t_i}^{t_{i-1}}) \quad (32) \end{aligned}$$

30

40

50

ただし、 e_i はラベル付き音響信号列 2011 - s の要素番号 i に対応する音響イベントラベルが表す音響イベント、 N_s はラベル付き音響信号列 2011 - s の要素数（要素番号 i の最大値）、 z_i は音響信号列 2011 a - s の要素番号 i に対応する音響信号が表す状況の候補、 t_{i-1} は音響信号列 2011 a - s の要素番号 i - 1 に対応する音響信号が表す状況の候補 t_{i-1} から要素番号 i に対応する音響信号が表す状況の候補 t_i への遷移確率を表す。また、 t は状況の候補 t が音響イベント（ただし、 $\{1, \dots, E\}$ ）を生成する確率 $P(t | s)$ を s 番目の要素とする E 次元ベクトルを表す。 t_s は音響信号列 2011 a - s が状況の候補 t（ただし、 $t \in \{1, \dots, T\}$ ）を生成する確率 $P(t | s)$ を t 番目の要素とする T 次元ベクトルを表す。

10

【0228】

[生成モデルの算出過程の説明]

モデル化部 2102 は、入力されたラベル付き音響信号列 2011 から、音響信号列が状況の候補を生成する確率 $P(\text{状況の候補} | \text{音響信号列})$ を算出し、それを音響信号 - 状況生成モデル 12 として記憶部 2103 に格納する。モデル化部 2102 は、状況の候補が音響イベントを生成する確率 $P(\text{音響イベント} | \text{状況の候補})$ を算出し、それを状況 - 音響イベント生成モデル 13 として記憶部 2103 に格納してもよい。また、モデル化部 2102 は、状況の候補の遷移確率を算出し、それを状況遷移モデル 2017 として記憶部 2103 に格納してもよい。

20

【0229】

さらにモデル化部 2102 は、入力されたラベル付き音響信号列 2011 に対して尤もらしい状況を表す状況ラベル 19 を選択して外部に出力したり、記憶部 2103 に格納したりしてもよい。選択される状況ラベル 19 は、各時間区間にに対して 1 個でもよいし、複数個でもよい。また、各時間区間ににおいて尤もらしさが閾値を超えたすべての状況を表すラベルを状況ラベル 19 の列を外部に出力したり、記憶部 2103 に格納したりしてもよい。

【0230】

音響信号 - 状況生成モデル 12 および状況 - 音響イベント生成モデル 13 の算出は、入力されたラベル付き音響信号列 2011 の音響イベントの時系列の並びにおける、状況の候補に対応する音響イベントの組み合わせと状況の候補の時系列の並びとの同時事後確率を最大化する処理を用いて行われる。例えば、マルコフ連鎖モンテカルロ法（MCMC 法 : Markov Chain Monte Carlo methods）や変分ベイズ法（VB 法 : Variational Bayes methods）などの手法を用いることができる。MCMC 法には、M-H アルゴリズムやギブスサンプリングなどの手法がある。ここでは一例としてギブスサンプリングによる生成モデルの算出手法を説明する。

30

【0231】

[生成モデルの算出方法の例示]

(i) モデル化部 2102 は、ハイパーパラメータとしてパラメータ α, β および γ を事前に決定する。

(ii) モデル化部 2102 は、全ての音響信号列 2011 a - s（ただし $s = 1, \dots, S$ ）の全ての要素番号 i に対応する要素に状況の候補 t_i をランダムに割り当てる。また、要素番号 i に対応する要素への状況の候補 t_i の割り当てには、事前にモデル化部 2102 に設定しておいた割り当てであってもよい。

40

【0232】

(iii) モデル化部 2102 は、全ての音響信号列 2011 a - s（ただし $s = 1, \dots, S$ ）の各要素番号 i（要素番号 $i - 1$ または $i + 1$ に対応する要素が存在しない要素番号 i を除く）について、以下の(iii-1)および(iii-2)を交互に、規定の回数（正值であり、例えば 1 ~ 3000 回程度）、もしくは、所望の結果が得られるまで（例えば、割り当ての前後において、状況の割り当て先の変化が一定の閾値（例えば 0.3%）以下になるまでなど）繰り返す。

50

【0233】

(iii-1) モデル化部 2102 は、以下の式により、更新対象のラベル付き音響信号列の要素番号 i を除いた要素番号に対応する音響信号列に対して、要素番号 i の要素に状況の候補 z_i に割り当てられる確率分布 $P(z_i | z_{-i}, e_{-i}, \cdot, \cdot, \cdot, \cdot)$ を更新する。

【数31】

$$P(z_i | z_{-i}, e_{-i}, \alpha', \beta', \gamma')$$

$$\begin{aligned} &\propto \frac{C_{et}^{ET} + \beta'}{\left(\sum_{e=1}^E C_{et}^{ET}\right) + E\beta'} \times \frac{C_{ts}^{TS} + \alpha'}{\left(\sum_{t=1}^T C_{ts}^{TS}\right) + T\alpha'} \\ &\times \frac{(R_{t_i}^{t_{i-1}} + \gamma)(R_{t_{i+1}}^{t_i} + I(t_{i-1} = t_i) \times I(t_i = t_{i+1}) + \gamma')}{\left(\sum_{t_{i+1} \in \{1, \dots, T\}} R_{t_{i+1}}^{t_i}\right) + I(t_{i-1} = t_i) + T\gamma'} \quad (33) \end{aligned}$$

ただし、 C_{t}^{ET} は更新対象のラベル付き音響信号列（初期値はラベル付き音響信号列 2011）で音響イベント t に状況 t が割り当てられた回数を表し、 C_{ts}^{TS} は更新対象のラベル付き音響信号列で状況 t が更新対象の音響信号列（初期値は音響信号列 2011 a - s）に割り当てられた回数を表す。なお、表記制約上の都合から「 C_{t}^{ET} 」「 C_{ts}^{TS} 」と表記するが、本来は式(33)に示すように「 C_{t}^{ET} 」の「ET」は「 t 」の上に表記され、「 C_{ts}^{TS} 」の「TS」は「 ts 」の上に表記される。 z_{-i} は要素番号 i 以外の要素番号に対応する状況からなる列、 e_{-i} は要素番号 i 以外の要素番号に対応する音響イベントからなる列を表す。また、 $t_{i-1} = t_i$ のときに $I(t_{i-1} = t_i) = 1$ 、 $t_{i-1} \neq t_i$ のときに $I(t_{i-1} = t_i) = 0$ となり、 $t_i = t_{i+1}$ のときに $I(t_i = t_{i+1}) = 1$ 、 $t_i \neq t_{i+1}$ のときに $I(t_i = t_{i+1}) = 0$ となる。

【0234】

(iii-2) モデル化部 2102 は、上記の更新式(33)で得られた確率分布 $P(z_i | z_{-i}, e_{-i}, \cdot, \cdot, \cdot, \cdot)$ に従って、各要素番号 i の要素に割り当てる状況の候補 z_i をサンプリング（選択）する。モデル化部 2102 は、このようにサンプリングした状況の候補 z_i を要素番号 i の要素に割り当て、要素番号 i に対応する状況の候補を更新する。これにより更新対象のラベル付き音響信号列を更新する。

【0235】

(iv) モデル化部 2102 は、上記の(iii-1)(iii-2)の繰り返しによって最終的に得られた C_{t}^{ET} 、 C_{ts}^{TS} および $R_{t_i}^{t_{i-1} t_{i+1}}$ を用い、以下を計算する。

【数32】

$$\phi_{et} = \frac{C_{et}^{ET} + \beta'}{\left(\sum_{e=1}^E C_{et}^{ET}\right) + E\beta'} \quad (34)$$

$$\theta_{ts} = \frac{C_{ts}^{TS} + \alpha'}{\left(\sum_{t=1}^T C_{ts}^{TS}\right) + T\alpha'} \quad (35)$$

$$\pi_{t_{i+1}}^{t_i} = \frac{(R_{t_i}^{t_{i-1}} + \gamma')(R_{t_{i+1}}^{t_i} + \gamma')}{\left\{\sum_{t_i \in \{1, \dots, T\}} R_{t_i}^{t_{i-1}}\right\} \left\{\sum_{t_{i+1} \in \{1, \dots, T\}} R_{t_{i+1}}^{t_i}\right\} + T\gamma'} \quad (36)$$

【0236】

これによってモデル化部 2102 は、状況が音響イベントを生成する確率 π_t （ただし、 $t \in \{1, \dots, T\}$ 、 $\{1, \dots, E\}$ ）の集合を得、それを状況 - 音響

10

20

30

40

50

イベント生成モデル 13 とする。例えばモデル化部 2102 は、確率 t_s を (t, s) 要素とする $T \times E$ 行列を状況 - 音響イベント生成モデル 13 とする。また、モデル化部 2102 は、音響信号列が状況を生成する確率 t_{s_i} (ただし、 $s_i \in \{1, \dots, S\}$)、 $t_{s_i} \in \{1, \dots, T\}$ の集合を得、それを音響信号 - 状況生成モデル 12 とする。例えばモデル化部 2102 は、確率 t_{s_i} を (s_i, t) 要素とする $S \times T$ 行列を音響信号 - 状況生成モデル 12 とする。また、モデル化部 2102 は、これによって状況の候補の遷移確率 $t_{s_i, t_{i+1}}$ (ただし、 $t_i, t_{i+1} \in \{1, \dots, T\}$) の集合を得、それを状況遷移モデル 2017 とする。例えばモデル化部 2102 は、 $t_i, t_{i+1} \in \{1, \dots, T\}$ に対応する遷移確率 $t_{s_i, t_{i+1}}$ を (t_i, t_{i+1}) 要素とする $T \times T$ 行列を状況遷移モデル 2017 とする。

10

【0237】

或いは、モデル化部 2102 は、(iii-1)の繰り返し処理時に、式(33)の算出過程で確率 $t_s, t_{s_i}, t_{s_i, t_{i+1}}$ をそれぞれ 1 個以上サンプリング(計算)し、式(34)(35)(36)に代えて、サンプリングされた確率 t_s の平均値、確率 t_{s_i} の平均値、および $t_{s_i, t_{i+1}}$ の平均値を用い、それぞれ、状況 - 音響イベント生成モデル 13、音響信号 - 状況生成モデル 2012、および状況遷移モデル 2017 を得てもよい。

【0238】

また、上記更新過程でサンプリング(選択)し、各要素番号 i の要素に割り当てた状況の候補 z_i を分析することで、各ラベル付き音響信号列 2011-s の音響信号列がどの状況によって生成されたかを知ることができる。モデル化部 1102 は、このように各要素番号 i の要素に割り当てた状況の候補 z_i を表すラベル、またはそれらの一部を状況ラベル 19 として出力してもよい。

20

【0239】

本実施例では、音響信号が状況を生成する確率や状況が音響イベントを生成する確率のみでなく、状況の時間遷移確率も考慮して学習処理を行うことで、遷移確率によって特徴付けられる状況を精度よくモデル化できる。また、遷移確率によって特徴付けられる状況を精度よく分析できる。

【0240】

<実施例(3)-1-2:図25>

30

[音響信号列から音響信号 - 状況生成モデル、状況 - 音響イベント生成モデルおよび状況遷移モデルを算出し、状況を分析]

本実施例では、音響信号列 2015-1, ..., 2015-S を入力とし、状況の時間変化を考慮に入れ、音響信号 - 状況生成モデル 2012、状況 - 音響イベント生成モデル 13、状況遷移モデル 2017 の少なくとも何れかを得る。すなわち、本実施例では、音響信号から音響特徴量を得、得られた音響特徴量を用いて音響イベントを決定し、それによって音響イベントラベル列を得てから学習処理を行い、音響信号 - 状況生成モデル 2012、状況 - 音響イベント生成モデル 13、状況遷移モデル 2017 の少なくとも何れかを得る。また状況の分析を行ってもよい。

【0241】

40

図 25 に例示するように、本実施例のモデル処理装置 2110 は、特徴量算出部 111、音響イベント判定部 112、音響イベントモデルデータベース(DB) 113、音響信号列合成部 2101、モデル化部 2102、および記憶部 2103 を有する。モデル処理装置 2110 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0242】

まず特徴量算出部 111 に音響信号列 2015-1, ..., 2015-S が入力される。各音響信号列 2015-s (ただし、 $s \in \{1, \dots, S\}$) は、前述の時間区間にごとに区分された要素からなり、各要素には要素番号が付されている。

【0243】

50

特徴量算出部 111 は、各音響信号列 2015 - s から、それぞれに対応する音響特徴量列（ベクトル）を算出して出力する。音響特徴量列の具体例は、実施例（1）-1-2 と同じである。

【0244】

音響イベント判定部 112 は、各要素番号 i に対して決定した音響イベントを表す音響イベントラベルを、音響信号列 2015 - s の各要素番号 i の要素に付与する。音響イベント判定部 112 は、この処理を入力された音響信号列 2015 - 1, ..., 2015 - S のすべての要素（すべての要素番号 i）について行い、その結果得られる音響イベントラベル付き音響信号列 2011 - 1, ..., 2011 - S を出力する。

【0245】

音響イベントラベル付き音響信号列 2011 - 1, ..., 2011 - S は、音響信号列合成部 2101 に入力される。以降の処理は実施例（3）-1-1 と同じである。

【0246】

なお、音響信号列合成部 2101 で音響イベントラベル付き音響信号列 2011 - 1, ..., 2011 - S の合成処理を行うことに代えて、特徴量算出部 111 の前段で音響信号列 2015 - 1, ..., 2015 - S の合成処理を行っても良いし、音響イベント判定部 112 の前段で音響特徴量列の合成処理を行っても良い。

【0247】

<実施例（3）-1-3：図26>

[音響特徴量列から音響信号・状況生成モデル、状況・音響イベント生成モデルおよび状況遷移モデルを算出し、状況を分析]

本実施例では、音響特徴量列 2016 - 1, ..., 2016 - S を入力とし、状況の時間変化を考慮に入れ、音響信号・状況生成モデル 2012、状況・音響イベント生成モデル 13、状況遷移モデル 2017 の少なくとも何れかを得る。すなわち、本実施例では、音響信号から音響特徴量を得、得られた音響特徴量を用いて音響イベントを決定し、それによって音響イベントラベル列を得てから学習処理を行い、音響信号・状況生成モデル 2012、状況・音響イベント生成モデル 13、状況遷移モデル 2017 の少なくとも何れかを得る。また状況の分析を行ってもよい。

【0248】

図26に例示するように、本実施例のモデル処理装置 2120 は、音響イベント判定部 112、音響イベントモデルデータベース（DB）113、音響信号列合成部 2101、モデル化部 2102、および記憶部 2103 を有する。モデル処理装置 2120 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【0249】

まず音響イベント判定部 112 に、音響特徴量列 2016 - 1, ..., 2016 - S が入力される。音響イベント判定部 112 は、実施例（1）-1-2 で例示した方法等によって、音響特徴量列 2016 - 1, ..., 2016 - S を用いて各要素番号 i に対して音響イベントを決定する。音響イベント判定部 112 は、各要素番号 i に対して決定した音響イベントを表す音響イベントラベルを、音響特徴量列 2016 - s の各要素番号 i の要素に付与する。音響イベント判定部 112 は、この処理を音響特徴量列 2016 - 1, ..., 2016 - S のすべての要素（すべての要素番号 i）について行い、その結果得られる音響イベントラベル付き音響信号列 2011 - 1, ..., 2011 - S を出力する。

【0250】

音響イベントラベル付き音響信号列 2011 - 1, ..., 2011 - S は、音響信号列合成部 2101 に入力される。以降の処理は実施例（3）-1-1 と同じである。

【0251】

なお、音響信号列合成部 2101 で合成処理を行うことに代えて、音響イベント判定部 112 の前段で音響特徴量列 2016 - 1, ..., 2016 - S の合成処理を行っても

10

20

30

40

50

良い。

【0252】

<実施例(3)-2-1:図27>

[音響イベントラベル付き音響信号列から状況を推定]

本実施例では、実施例(3)-1-1～(3)-1-3で説明したように得られた状況-音響イベント生成モデル13および/または状況遷移モデルを用い、新たに入力された音響イベントラベル付き音響信号列から状況を推定する。

【0253】

図27に例示するように、本実施例のモデル処理装置2200は、記憶部2103および生成モデル比較部2201を有する。モデル処理装置2200は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。10

【0254】

まず生成モデル比較部2201に音響イベントラベル付き音響信号列2021が入力される。音響イベントラベル付き音響信号列2021は、前述の時間区間ごとに区分された時系列の音響信号列、当該時間区間ごとに区分された音響信号列の各要素に対応する要素番号、および当該時間区間ごとに決定されて付与された音響イベントラベルを含む。要素番号および音響イベントラベルは、音響信号列の要素ごとに付与される。

【0255】

生成モデル比較部2201は、入力された音響イベントラベル付き音響信号列2021と、記憶部2103に格納された状況-音響イベント生成モデル13および/または状況遷移モデルとを比較し、音響イベントラベル付き音響信号列2021に対し、最も適切であると判断した状況、又は最も適切なものから順番に複数個の状況を決定し、それらを推定結果として出力する。具体的には、例えば、実施例(2)-2-1の生成モデル比較部1201と同じ方法で、音響イベントラベル付き音響信号列2021に対応する状況の推定を行うことができる。入力された音響イベントラベル付き音響信号列2021の音響イベントの分布と、状況の候補に対応する確率P(音響イベント|状況の候補)で表される音響イベントの分布との近さに基づいて、状況が推定されてもよい。また、入力された音響イベントラベル付き音響信号列2021の音響イベントの時系列の並びに対する、確率P(音響イベント|状況の候補)のもとでの、状況の候補の尤もらしさに基づいて、状況が推定されてもよい。20

【0256】

<実施例(3)-2-2:図28>

[音響信号列から状況を推定]

本実施例では、実施例(3)-1-1～(3)-1-3で説明したように得られた状況-音響イベント生成モデル13および/または状況遷移モデルを用い、新たに入力された音響信号列から状況を推定する。

【0257】

図28に例示するように、本実施例のモデル処理装置2210は、特徴量算出部211、音響イベント判定部212、音響イベントモデルDB113、記憶部2103、生成モデル比較部2201を有する。モデル処理装置2210は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。40

【0258】

まず特徴量算出部211に音響信号列22が入力される。特徴量算出部211は、音響信号列22から音響特徴量列(ベクトル)を算出して出力する。例えば特徴量算出部211は、前述した特徴量算出部111と同じ方法で音響特徴量列を算出する。

【0259】

音響イベント判定部212は、前述した音響イベント判定部112と同じ方法で、特徴量算出部211から出力された音響特徴量列と、音響イベントモデルDB113に記憶されている複数の音響イベントモデルとをそれぞれ比較し、音響信号列22の全ての要素について音響イベントを決定する。音響イベント判定部212は、決定した音響イベントを50

表す音響イベントラベルを音響信号列 2 2 の各要素に付することで、音響イベントラベル付き音響信号列 2 0 2 1 を生成して出力する。

【 0 2 6 0 】

音響イベントラベル付き音響信号列 2 0 2 1 は、生成モデル比較部 2 2 0 1 に入力される。以降の処理は実施例(3)-2-1と同じである。

【 0 2 6 1 】

< 実施例(3)-2-3 : 図 2 9 >

[音響特徴量から状況を推定]

本実施例では、実施例(3)-1-1 ~ (3)-1-3 で説明したように得られた状況 - 音響イベント生成モデル 1 3 および/または状況遷移モデルを用い、新たに入力された音響特徴量列から状況を推定する。

10

【 0 2 6 2 】

図 2 9 に例示するように、本実施例のモデル処理装置 2 2 2 0 は、音響イベント判定部 2 1 2 、音響イベントモデル DB 1 1 3 、記憶部 2 1 0 3 、生成モデル比較部 2 2 0 1 を有する。モデル処理装置 2 2 2 0 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

【 0 2 6 3 】

まず音響イベント判定部 2 1 2 に音響特徴量列 2 3 が入力される。音響イベント判定部 2 1 2 は、前述した音響イベント判定部 1 1 2 と同じ方法で、音響特徴量列 2 3 の音響特徴量列と、音響イベントモデル DB 1 1 3 に記憶されている複数の音響イベントモデルとをそれぞれ比較し、音響特徴量列 2 3 が含む全ての要素について音響イベントを決定する。音響イベント判定部 2 1 2 は、決定した各要素の音響イベントを表す音響イベントラベルを、音響特徴量列 2 3 が含む音響信号列に付することで、音響イベントラベル付き音響信号列 2 0 2 1 を生成して出力する。

20

【 0 2 6 4 】

音響イベントラベル付き音響信号列 2 0 2 1 は、生成モデル比較部 2 2 0 1 に入力される。以降の処理は実施例(3)-2-1と同じである。

【 0 2 6 5 】

< 実施例(3)-3-1 : 図 3 0 >

[音響イベントラベル付き音響信号列から状況を推定すると共に、状況 - 音響イベント生成モデル、音響信号 - 状況生成モデル、および状況遷移モデルを算出]

30

本実施例は実施例(3)-1-1 と実施例(3)-2-1 の組み合わせである。本実施例では、音響イベントラベル付き音響信号列 2 0 2 1 を入力として状況を推定することに加え、音響イベントラベル付き音響信号列 2 0 1 1 - 1 , . . . , 2 0 1 1 - S を入力とし、音響信号 - 状況生成モデル 1 2 、および状況 - 音響イベント生成モデル 1 3 の少なくとも何れかの算出も行う。さらに状況遷移モデル 2 0 1 7 や状況ラベル 1 9 の生成を行ってもよい。

【 0 2 6 6 】

図 3 0 に例示するように、本実施例のモデル処理装置 2 3 0 0 は、記憶部 2 1 0 3 , 2 3 0 3 、音響信号列合成部 3 0 1 、モデル化部 2 1 0 2 、および生成モデル比較部 2 2 0 1 を有する。モデル処理装置 2 3 0 0 は、例えば、汎用又は専用の公知のコンピュータに所定のプログラムが読み込まれることで構成される。

40

【 0 2 6 7 】

記憶部 2 3 0 3 には、前述した音響イベントラベル付き音響信号列 2 0 1 1 - 1 , . . . , 2 0 1 1 - S 、および前述した音響イベントラベル付き音響信号列 2 0 2 1 が格納されている。

【 0 2 6 8 】

音響イベントラベル付き音響信号列 2 0 1 1 - 1 , . . . , 2 0 1 1 - S , 2 0 2 1 は音響信号列合成部 3 0 1 に入力される。音響信号列合成部 3 0 1 は、音響イベントラベル付き音響信号列 2 0 1 1 - 1 , . . . , 2 0 1 1 - S , 2 0 2 1 を時系列方向につなぎ合

50

わせて一つの音響イベントラベル付き音響信号列を生成し、モデル化部 2102 に送出する。モデル化部 2102 は、実施例(3)-1-1 で説明したように、入力された音響イベントラベル付き音響信号列から、音響信号 - 状況生成モデル 2012、および状況 - 音響イベント生成モデル 13 を生成し、それらを記憶部 2103 に格納する。また、実施例(3)-1-1 で説明したように、モデル化部 2102 が状況遷移モデル 2017 を生成して記憶部 2103 に格納してもよいし、状況ラベル 19 を生成して記憶部 2103 に格納してもよい。

【0269】

音響イベントラベル付き音響信号列 2021 は生成モデル比較部 2201 に入力される。音響イベントラベル付き音響信号列 2021 は、実施例(3)-2-1 で説明したように、入力された音響イベントラベル付き音響信号列 2021 と、記憶部 2103 に格納された状況 - 音響イベント生成モデル 13 および/または状況遷移モデルとを比較し、音響イベントラベル付き音響信号列 2021 に対し、最も適切であると判断した状況、又は最も適切なものから順番に複数個の状況を選択し、それらを推定結果として出力する。

10

【0270】

さらに、音響イベントラベル付き音響信号列 2021 が、音響イベントラベル付き音響信号列 2011-1, ..., 2011-S とともに音響信号列合成部 301 に入力されてもよい。この場合、音響信号列合成部 301 は、入力されたすべての音響イベントラベル付き音響信号列 2011-1, ..., 2011-S, 2021 を時系列方向につなぎ合わせ(合成)、それによって一つの音響イベントラベル付き音響信号列 2011' (「第3音響イベントラベル列」)を得て出力する。モデル化部 1102 は、実施例(3)-1-1 のラベル付き音響信号列 2011 (「音響イベントラベル列」)に代えて、音響イベントラベル付き音響信号列 2011' (「第3音響イベントラベル列」)を用い、再度、実施例(3)-1-1 で説明したように、音響信号 - 状況生成モデル 2012、および状況 - 音響イベント生成モデル 13 の少なくとも何れかを生成し、それを記憶部 2103 に格納する。さらにモデル化部 1102 が、ラベル付き音響信号列 2011 に代えて、入力された音響イベントラベル付き音響信号列 2011' を用い、再度、実施例(3)-1-1 で説明したように、状況遷移モデル 2017 を生成して記憶部 2103 に格納してもよいし、状況ラベル 19 を生成して記憶部 2103 に格納してもよい。

20

【0271】

30

また、生成モデル比較部 2201 の処理およびモデル化部 2102 の処理のどちらを行っても良い。ただし、モデル化部 2102 の処理を行う前に生成モデル比較部 2201 の処理を行う場合、記憶部 2103 に予め得られた各生成モデルおよび/または状況遷移モデルが格納されている必要がある。

【0272】

40

また、音響イベントラベル付き音響信号列 2021 が、新たに入力された音響イベントラベル付き音響信号列とともに音響信号列合成部 301 に入力されてもよい。この場合、音響信号列合成部 301 がこれらを時系列方向につなぎ合わせ(合成)、それによって一つの音響イベントラベル付き音響信号列 (「第3音響イベントラベル列」)を得て、モデル化部 2102 に送出してもよい。

その他の処理は実施例(3)-1-1 および実施例(3)-2-1 と同様とする。

【0273】

<実施例(3)-3-2 : 図 31>

[音響信号列から状況を推定すると共に、状況 - 音響イベントの生成モデル、音響信号 - 状況生成モデル、および状況遷移モデルを算出]

本実施例は、実施例(3)-1-2 と実施例(3)-2-2 の組み合わせである。

本実施例では、音響信号列 2015-1, ..., 2015-S, 22 を入力として、学習によって、音響信号 - 状況生成モデル 12、および状況 - 音響イベント生成モデル 13 の少なくとも何れかを算出する。さらに状況遷移モデル 2017 や状況ラベル 19 の生成を行ってもよい。さらに本実施例では、状況 - 音響イベント生成モデル 13 および/ま

50

たは状況遷移モデルを用い、音響信号列22から状況を推定する。

【0274】

図31に例示するように、本実施例のモデル処理装置310は、特微量算出部111-1, ..., 111-S, 211、音響イベント判定部112-1, ..., 112-S, 212、音響イベントモデルDB113、および前述のモデル処理装置2300(図30参照)を有する。

【0275】

音響信号列2015-1, ..., 2015-Sは、それぞれ特微量算出部111-1, ..., 111-Sに入力される。特微量算出部111-1, ..., 111-Sは、実施例(1)-1-2で説明したように、音響信号列2015-1, ..., 2015-Sから、それぞれ音響特微量列を得て出力する。音響イベント判定部112-1, ..., 112-Sは、それぞれ、前述の音響イベント判定部112と同様に、入力された音響特微量列と音響イベントモデルDB113に記憶されている複数の音響イベントモデルとから、音響イベントラベル付き音響信号列2011-1, ..., 2011-Sを生成して出力する。

【0276】

音響信号列22は特微量算出部211に入力される。特微量算出部211は、実施例(1)-2-2で説明したように、音響信号列22から音響特微量列(ベクトル)を算出して出力する。音響イベント判定部212は、実施例(1)-2-2で説明したように、入力された音響特微量列と音響イベントモデルDB113に記憶されている複数の音響イベントモデルとから、音響イベントラベル付き音響信号列2021を生成して出力する。

【0277】

音響イベントラベル付き音響信号列2011-1, ..., 2011-S、2021は、記憶部2303(図30)に格納される。以降の処理は実施例(3)-3-1と同じである。

【0278】

<実施例(3)-3-3:図32>

[音響特微量から状況を推定すると共に、状況-音響イベント生成モデル、音響信号-状況生成モデル、および状況遷移モデルを算出]

本実施例は、実施例(3)-1-3と実施例(3)-2-3の組み合わせである。

本実施例では、音響特微量列2016-1, ..., 2016-S, 23を入力として、学習によって、音響信号-状況生成モデル12、および状況-音響イベント生成モデル13の少なくとも何れかを算出する。状況遷移モデル2017や状況ラベル19の生成を行ってもよい。さらに本実施例では、音響信号-状況生成モデル12、および状況-音響イベント生成モデル13を用い、音響特微量列23から状況を推定する。

【0279】

図32に例示するように、本実施例のモデル処理装置2320は、音響イベント判定部112-1, ..., 112-S, 212、音響イベントモデルDB113、および前述のモデル処理装置2300(図30参照)を有する。

【0280】

音響特微量列2016-1, ..., 2016-Sは、それぞれ音響イベント判定部112-1, ..., 112-Sに入力される。音響イベント判定部112-1, ..., 112-Sは、それぞれ、前述の音響イベント判定部112と同様に、入力された音響特微量列と音響イベントモデルDB113に記憶されている複数の音響イベントモデルとから、音響イベントラベル付き音響信号列2011-1, ..., 2011-Sを生成して出力する。

【0281】

音響特微量列23は音響イベント判定部212に入力される。音響イベント判定部212は、実施例(1)-2-3と同様に、入力された音響特微量列23の音響特微量列と音響イベントモデルDB113に記憶されている複数の音響イベントモデルとから、音響イ

10

20

30

40

50

ペントラベル付き音響信号列 2021 を生成して出力する。

【0282】

音響イベントラベル付き音響信号列 2011-1, ..., 2011-S、音響イベントラベル付き音響信号列 2021 は、記憶部 2303 (図 30) に格納される。以降の処理は実施例 (2)-3-1 と同じである。

【0283】

<変形例等>

なお、本発明は上述の実施例に限定されるものではない。例えば、モデル処理装置の処理が複数の装置で分散処理されてもよいし、上記の各実施例で記憶部や DB に格納された各データが複数の記憶部や DB に分散して格納されてもよい。例えば、行動 - 状況生成モデル、状況 - 音響イベント生成モデル、行動 - 音響イベント生成モデル、音響信号 - 状況生成モデル等の生成モデルや、行動遷移モデルや状況遷移モデル等が互いに異なる記憶部に格納されてもよい。

10

【0284】

音響信号列を使用しない場合 (例えば、実施例 (1)-1-1 等) には、音響イベントラベルおよび行動ラベル候補付き音響信号列に代えて、音響イベントラベルおよび行動ラベル候補列が用いられてもよい。同様に、音響信号列を使用しない場合 (例えば、実施例 (1)-3-1 等) には、音響イベントラベル付き音響信号列に代えて、音響イベントラベル列が用いられてもよい。また、音響イベントラベル列と行動ラベル候補列とが別々に入力されてもよい。

20

【0285】

音響イベントラベルおよび行動ラベル候補列の例は、音響イベントラベルおよび行動ラベル候補付き音響信号列から音響信号列を除いた列である。また、音響イベントラベル列の例は、音響イベントラベル付き音響信号列から音響信号列を除いた列である。

【0286】

また、ラベル列が時系列の順に入力され順次処理されるのであれば、音響イベントラベル付き音響信号列、音響イベントラベルおよび行動ラベル候補付き音響信号列等が要素番号を含まなくてもよい。

【0287】

また、各実施例のモデル処理装置が行動や状況の分析を行う場合には分析装置として機能する。各実施例のモデル処理装置が、生成モデルを出力することなく、行動や状況の分析のみを行い、その推定結果を出力してもよい。

30

【0288】

上述の各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的あるいは個別に実行されてもよい。その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能であることはいうまでもない。

【0289】

上述の構成をコンピュータによって実現する場合、各装置が有すべき機能の処理内容はプログラムによって記述される。このプログラムをコンピュータで実行することにより、上記処理機能がコンピュータ上で実現される。この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体の例は、非一時的な (non-transitory) 記録媒体である。このような記録媒体の例は、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等である。

40

【0290】

このプログラムの流通は、例えば、そのプログラムを記録した DVD、CD-ROM 等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

【0291】

50

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶装置に格納する。処理の実行時、このコンピュータは、自己の記憶装置に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆる A S P (Application Service Provider) 型のサービスによって、上述の処理を実行する構成としてもよい。

10

【 0 2 9 2 】

上記実施形態では、コンピュータ上で所定のプログラムを実行させて本装置の処理機能が実現されたが、これらの処理機能の少なくとも一部がハードウェアで実現されてもよい。

【 図 1 】

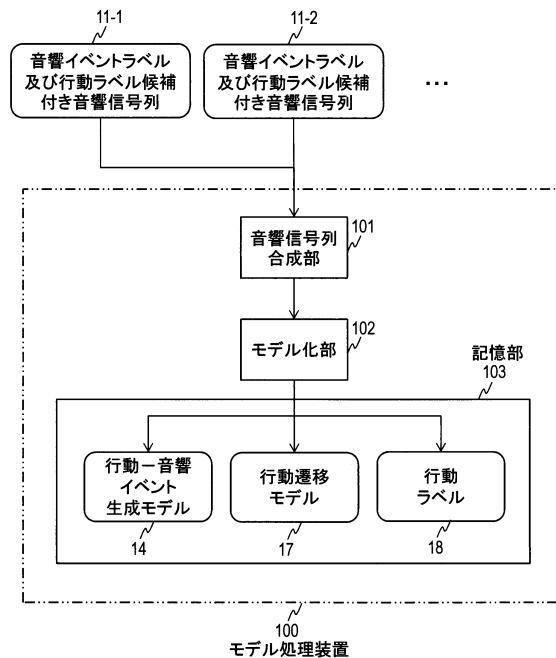


図1

【 図 2 】

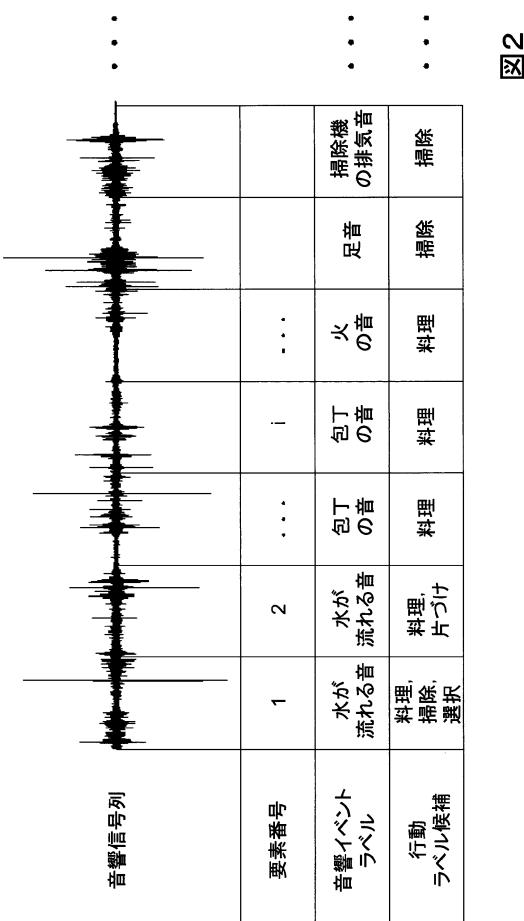
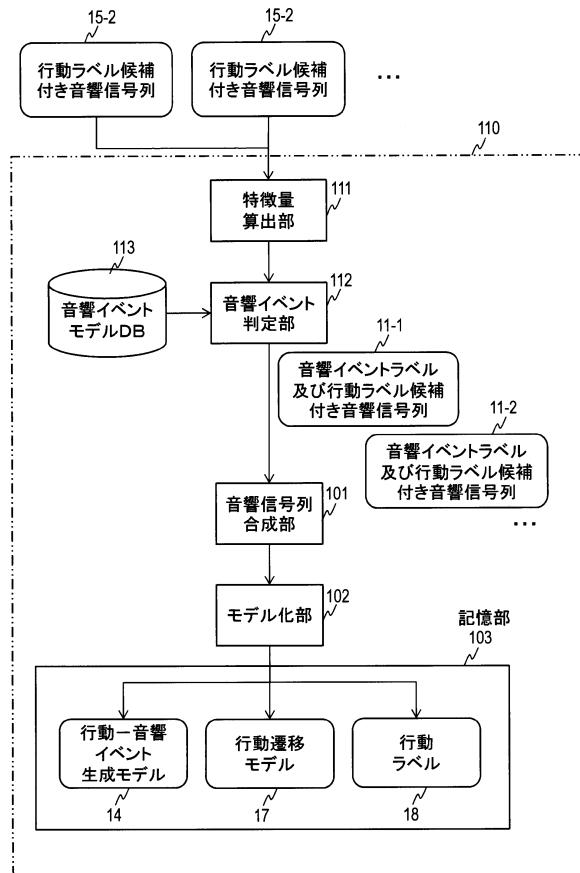


図2

【図3】



【図4】

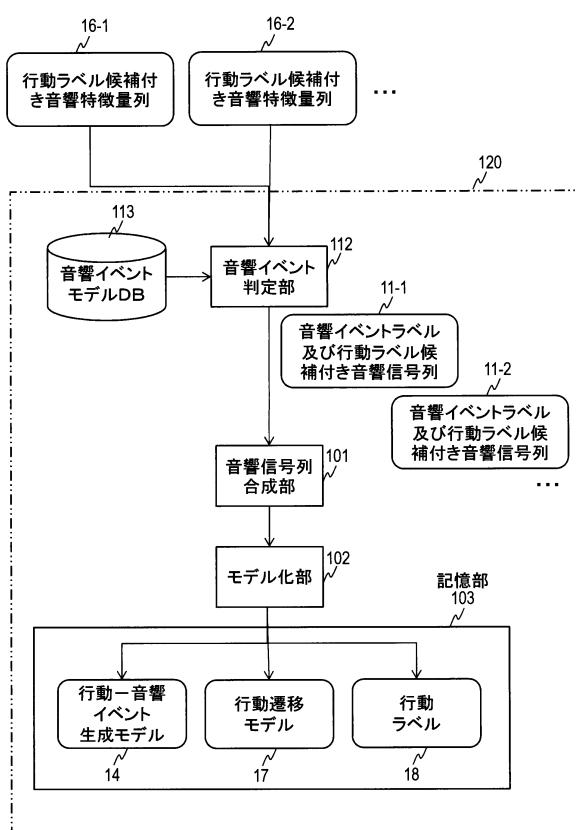


図4

【図5】

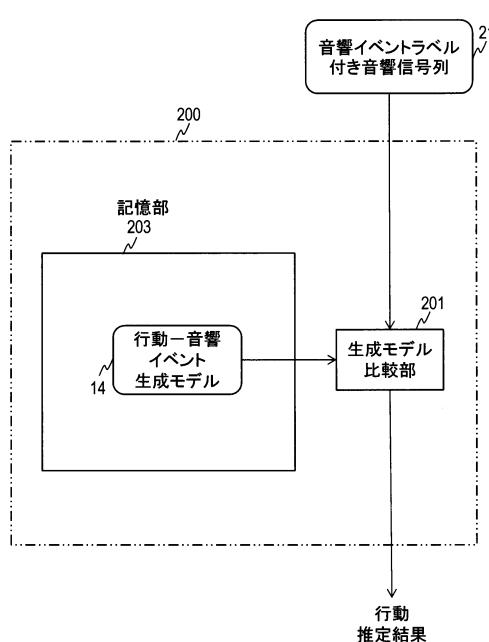


図5

【図6】

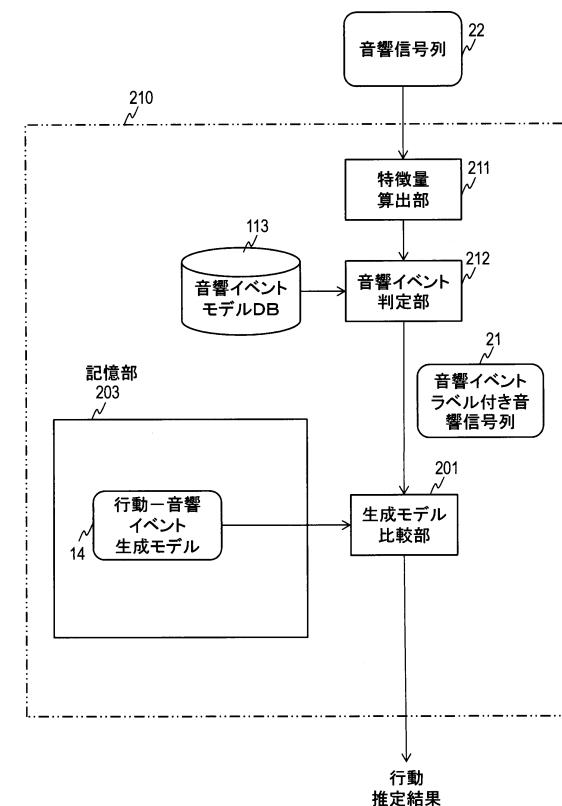


図6

【図7】

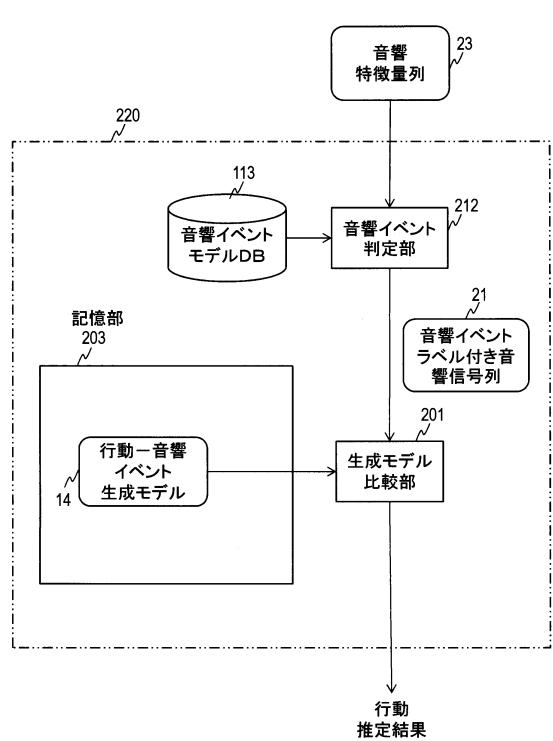


図7

【図8】

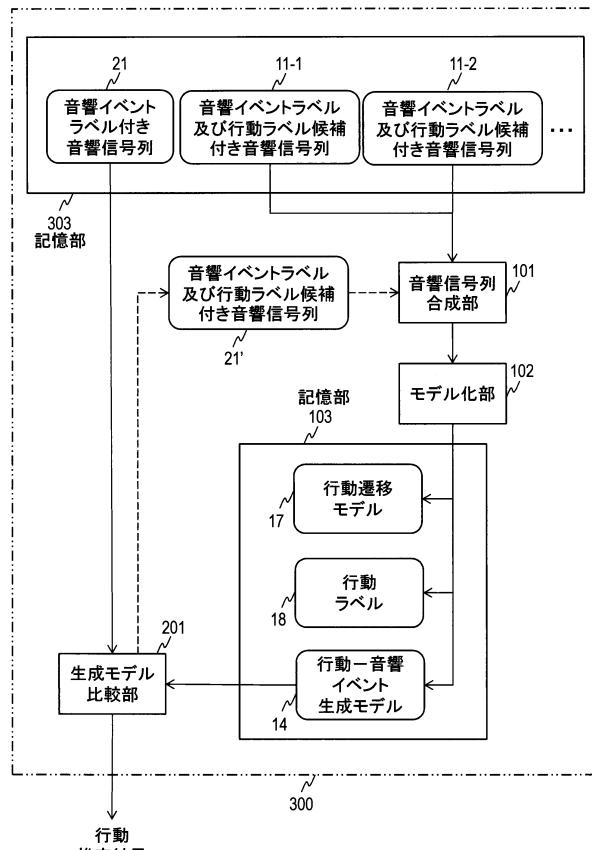


図8

【図9】

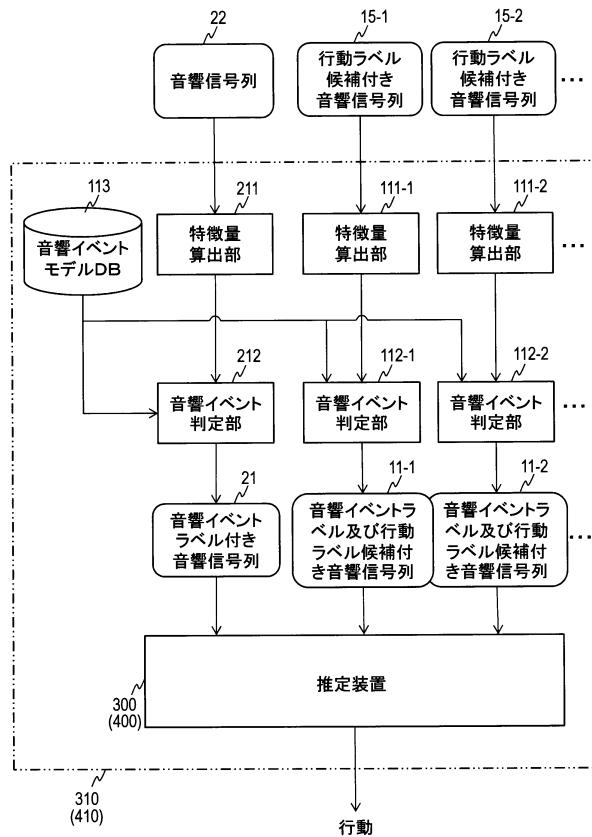


図9

【図10】

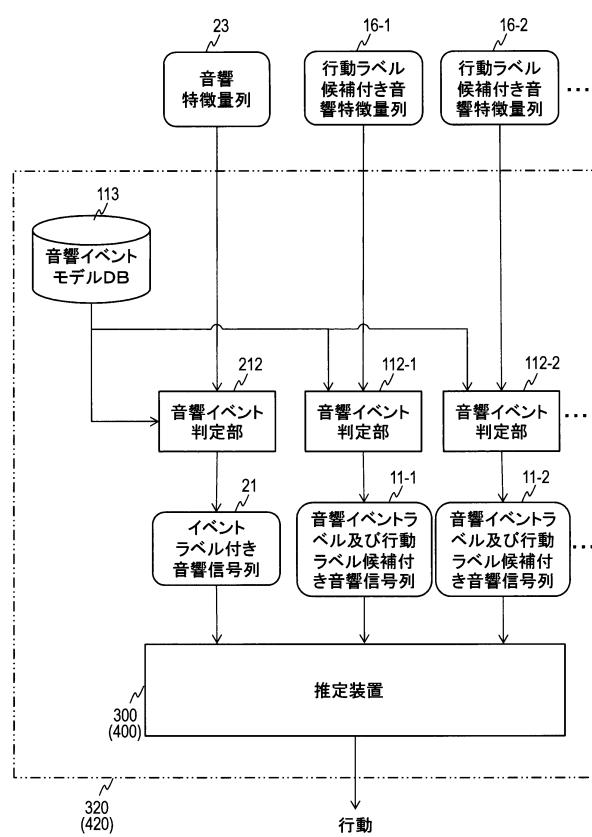


図10

【図 1 1】

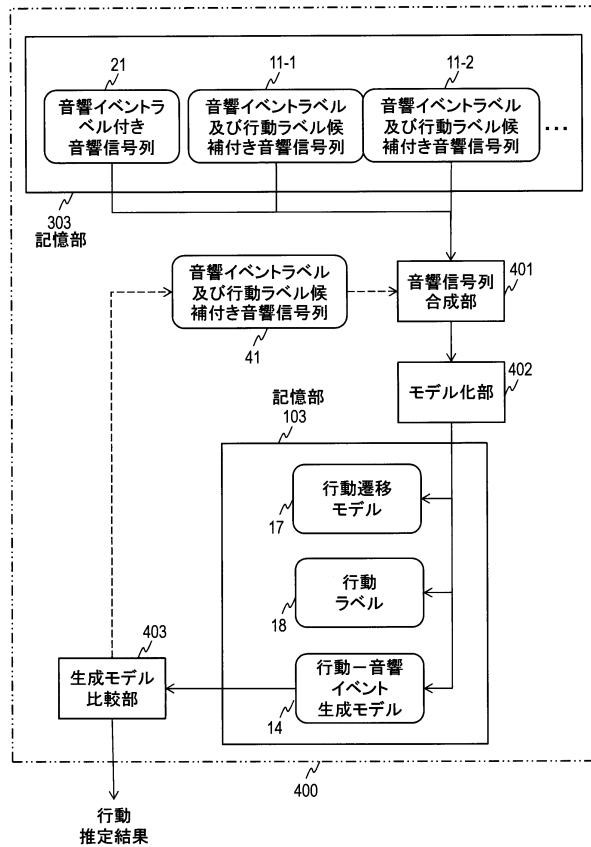


図 1 1

【図 1 2】

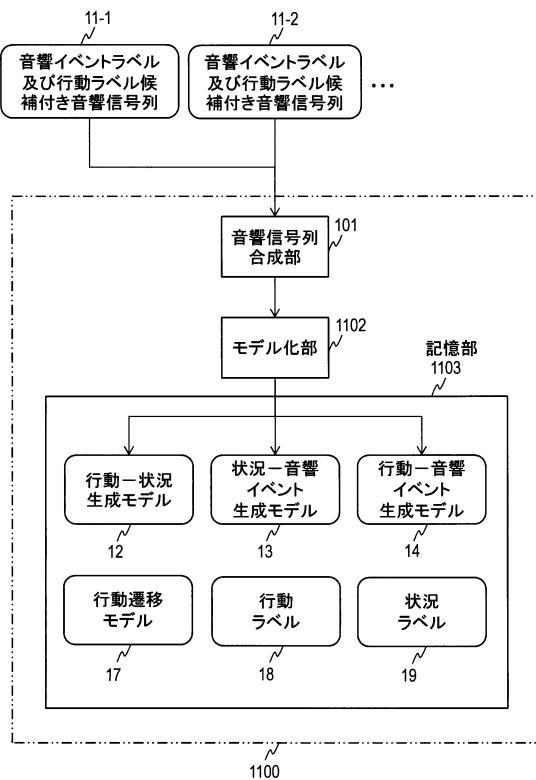


図 1 2

【図 1 3】

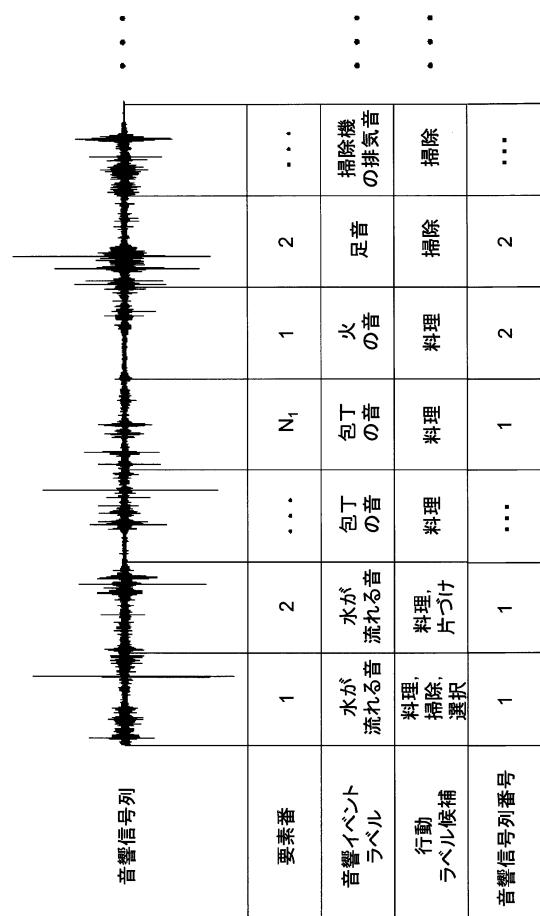


図 1 3

【図 1 4】

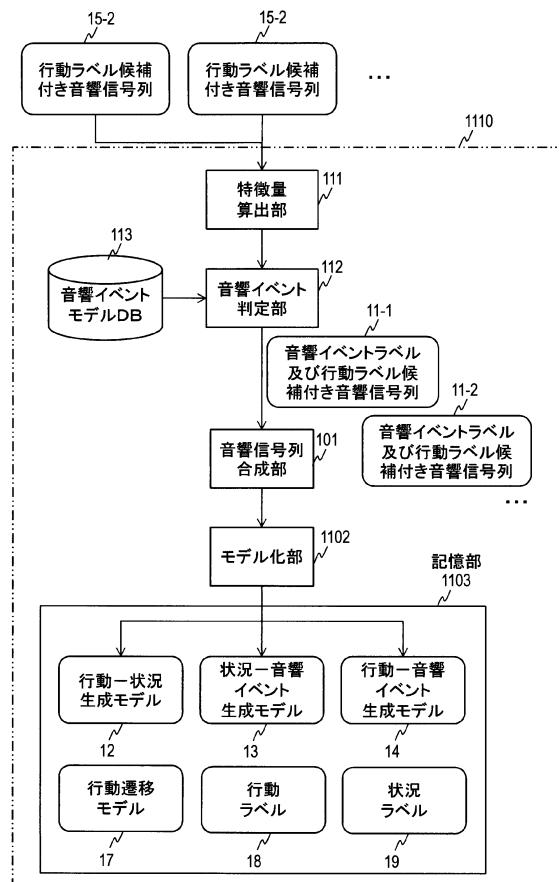


図 1 4

【図15】

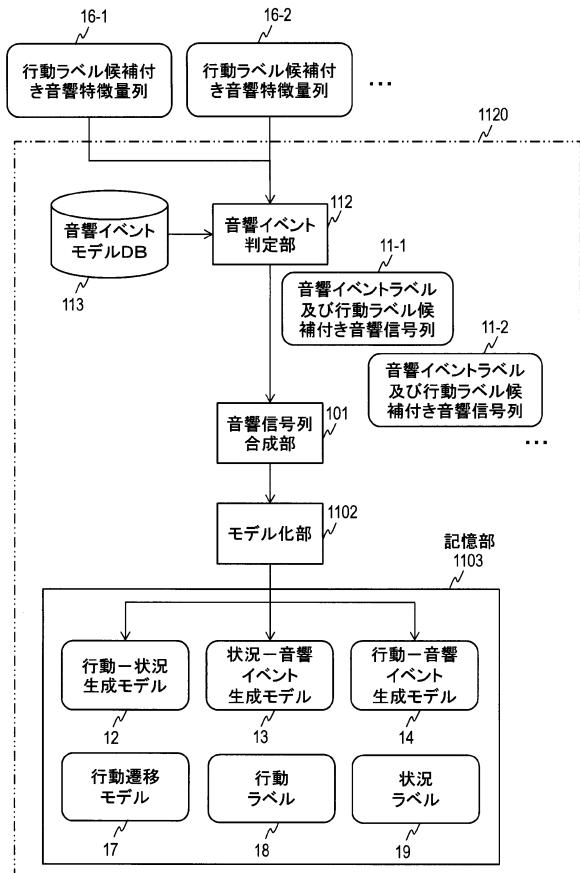


図15

【図16】

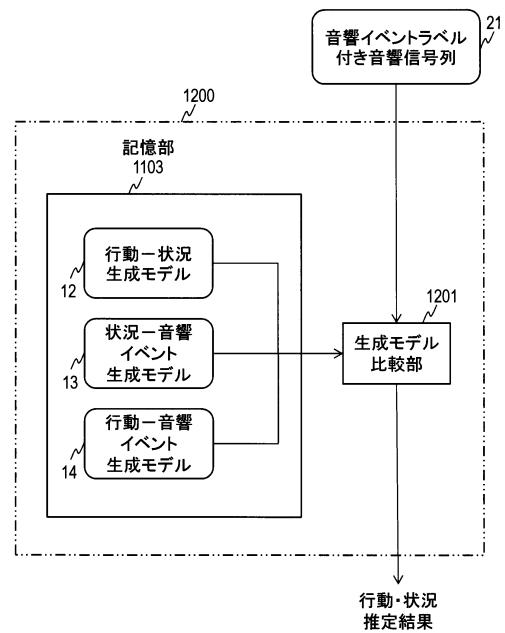


図16

【図17】

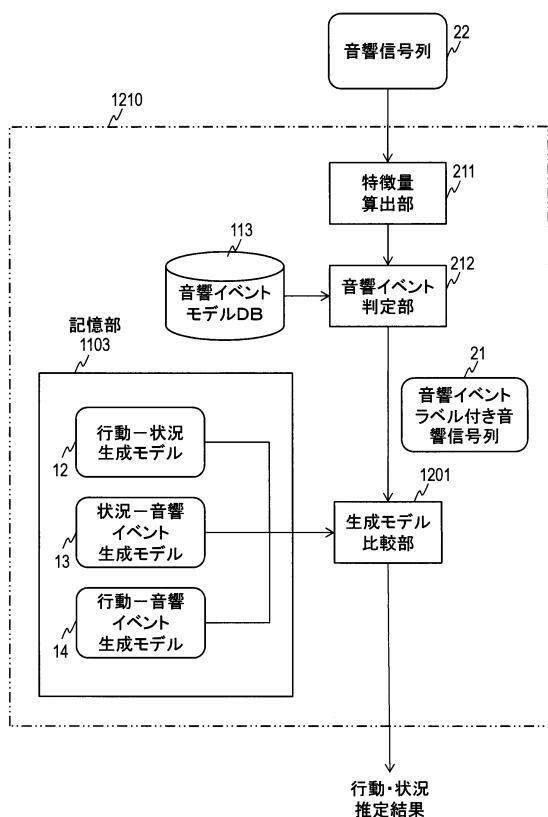


図17

【図18】

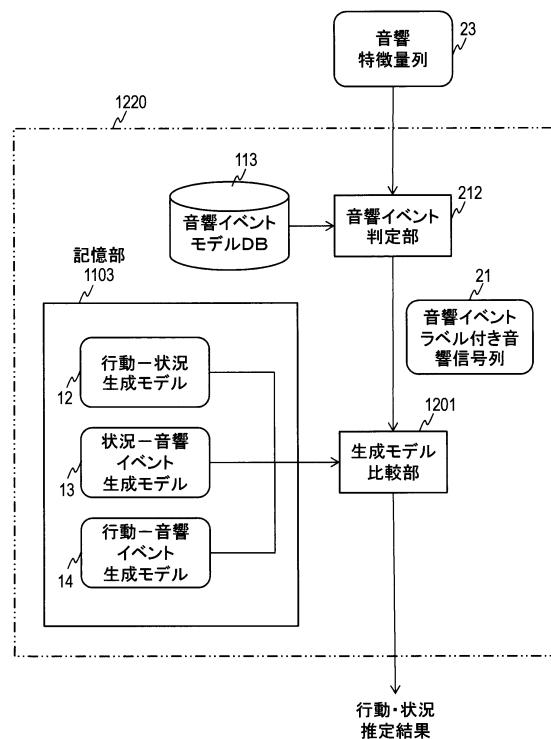


図18

【図19】

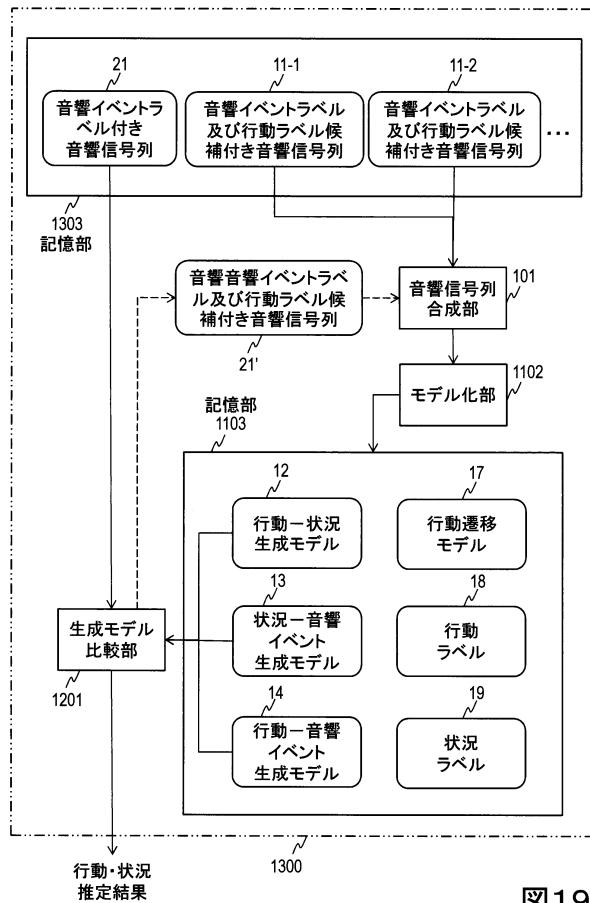


図19

【図20】

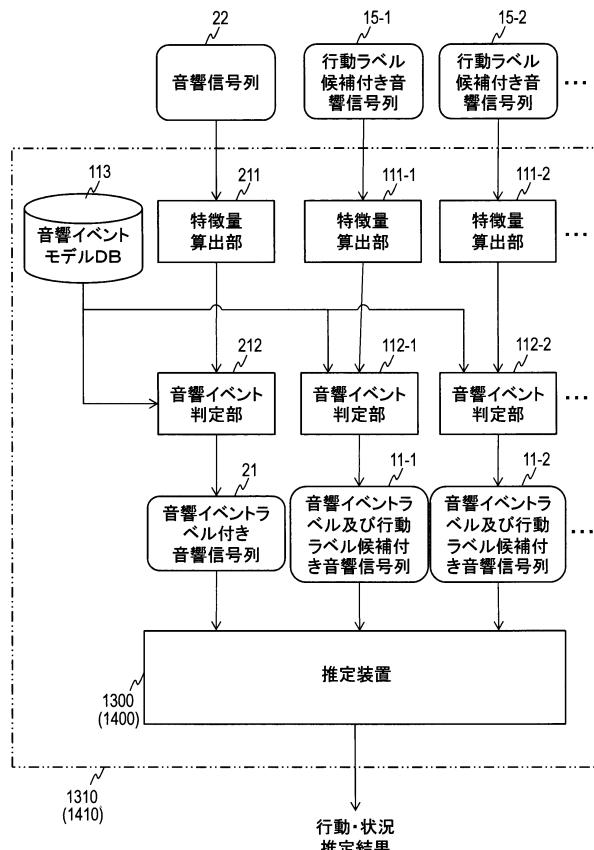


図20

【図21】

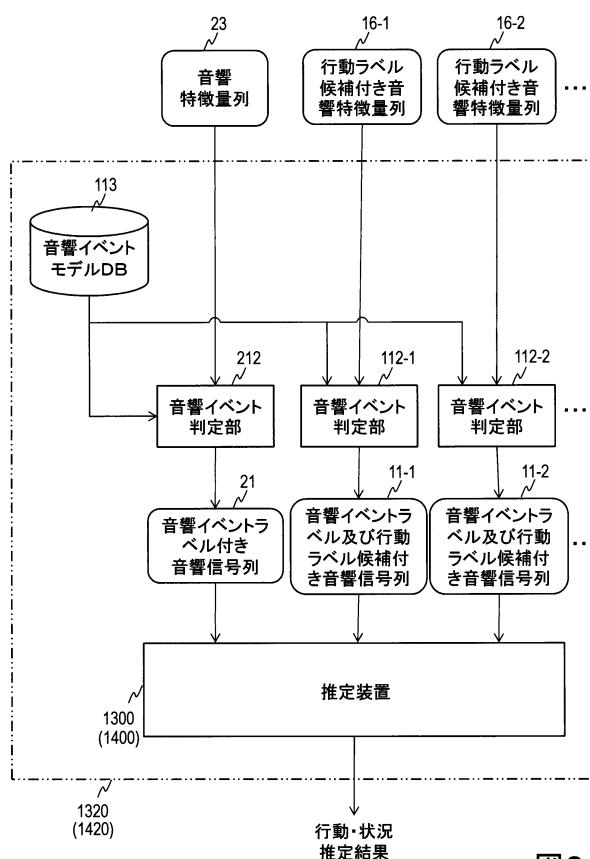


図21

【図22】

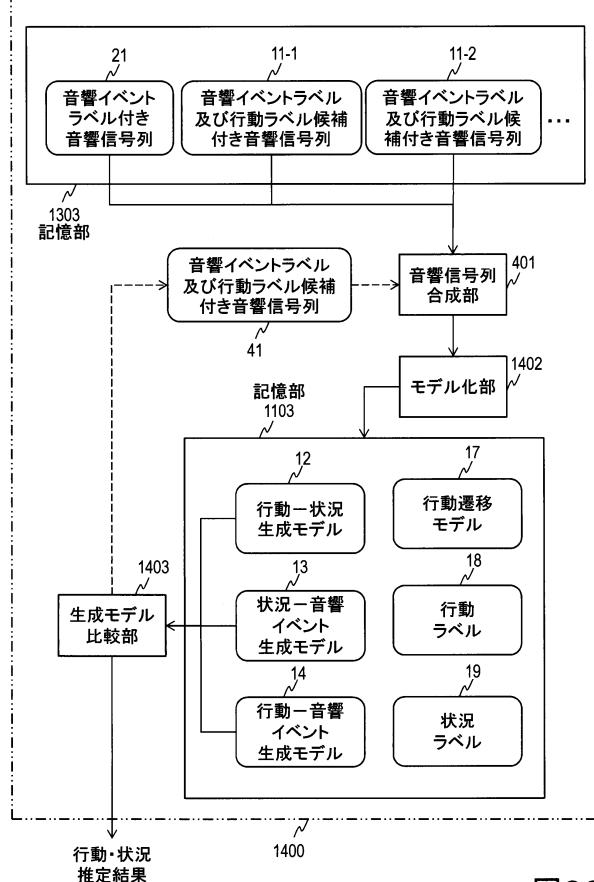


図22

【図23】

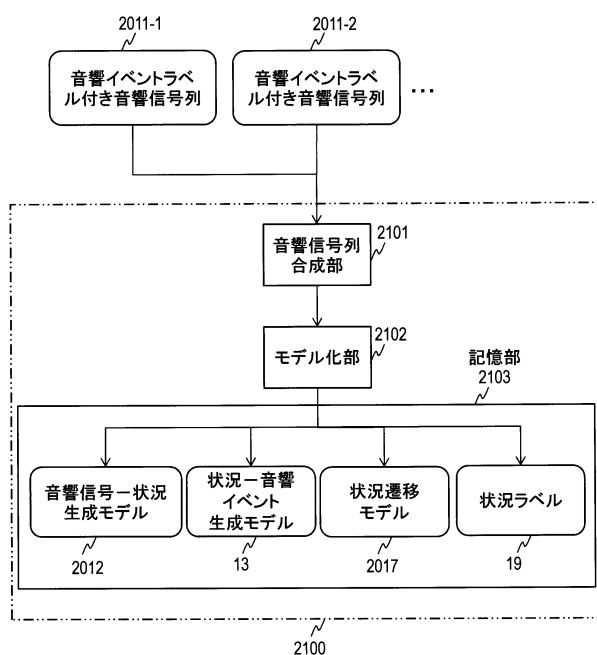
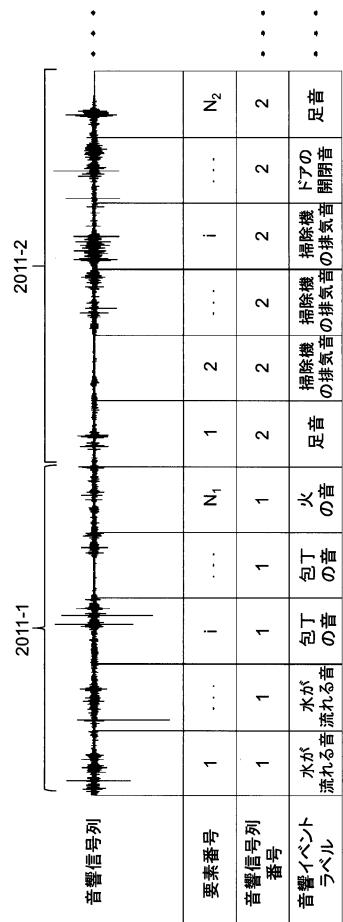


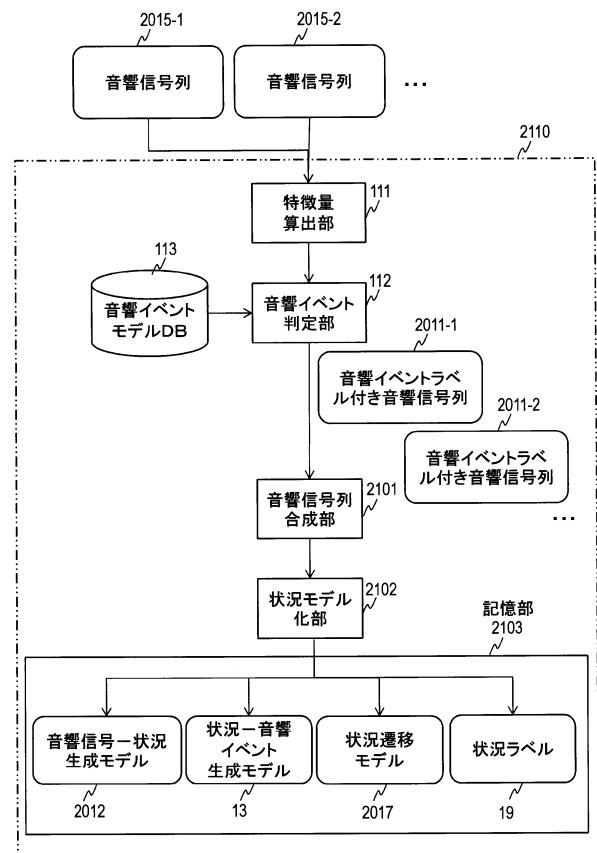
图23

【図24】



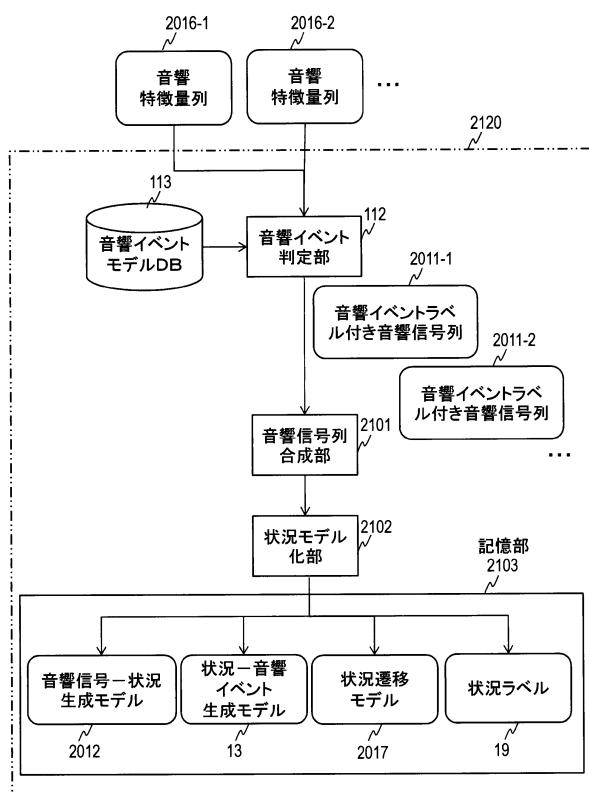
24

【図25】



25

【 図 2 6 】



26

【図27】

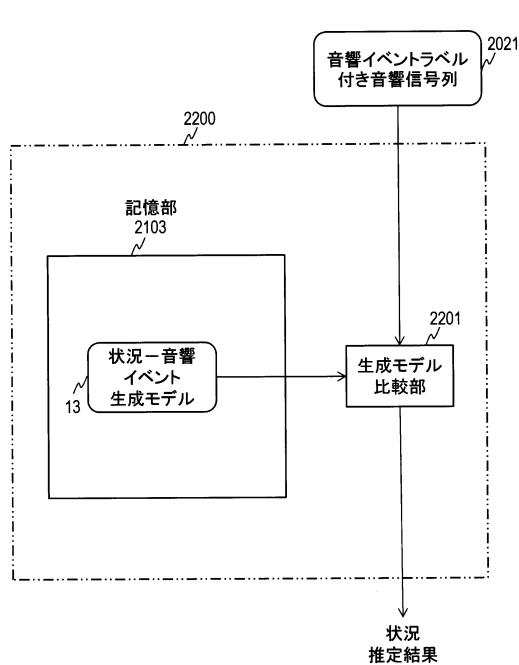


図27

【図28】

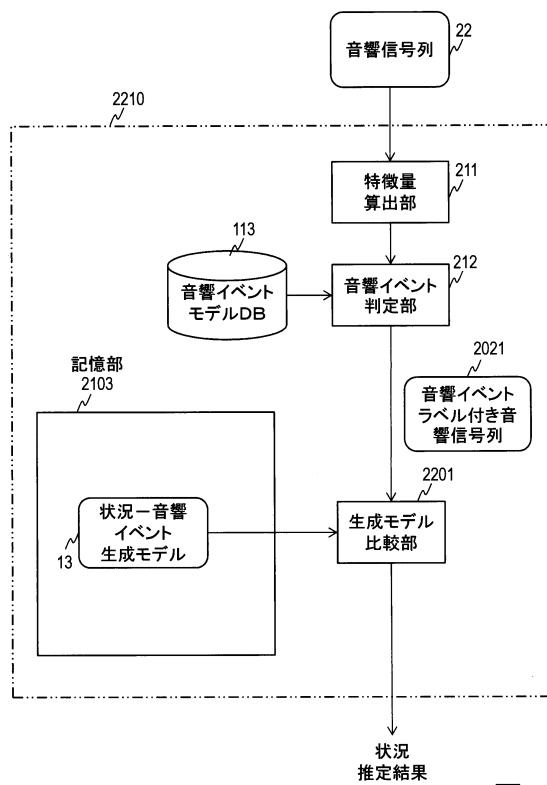


図28

【図29】

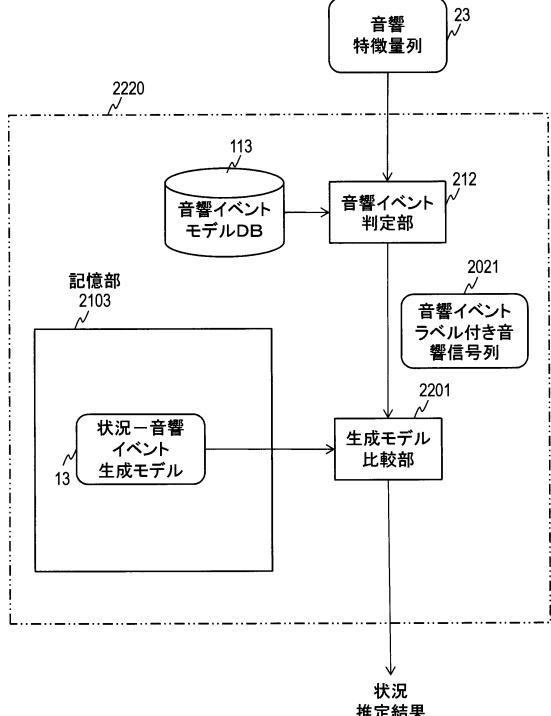


図29

【図30】

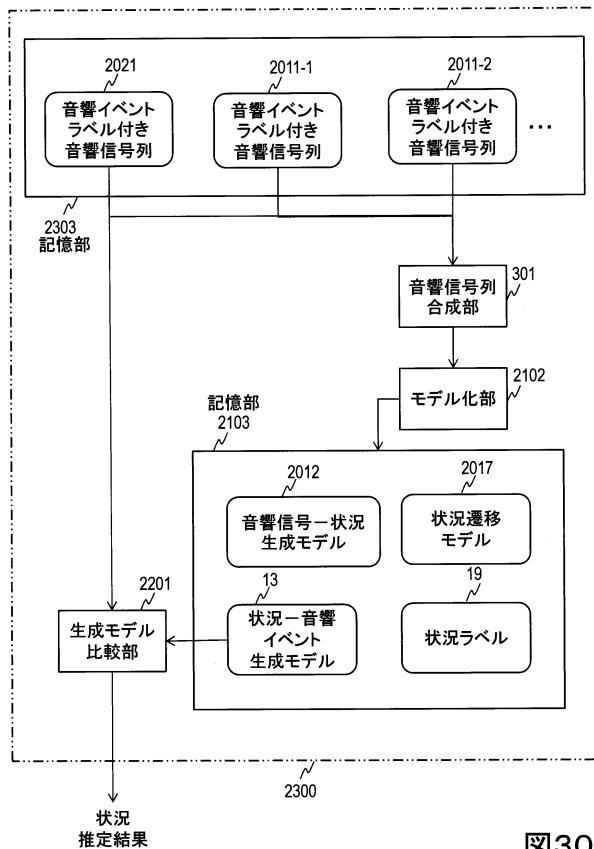
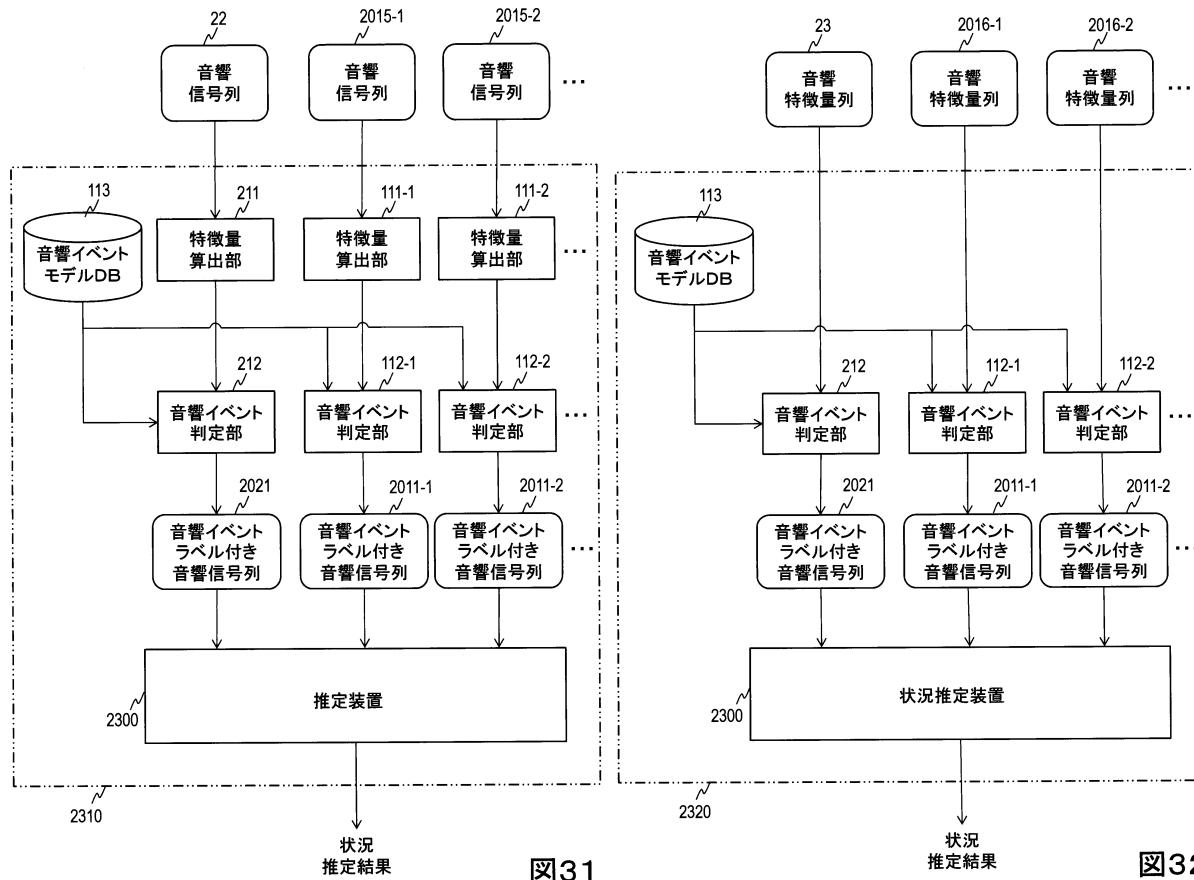


図30

【図31】



【図32】

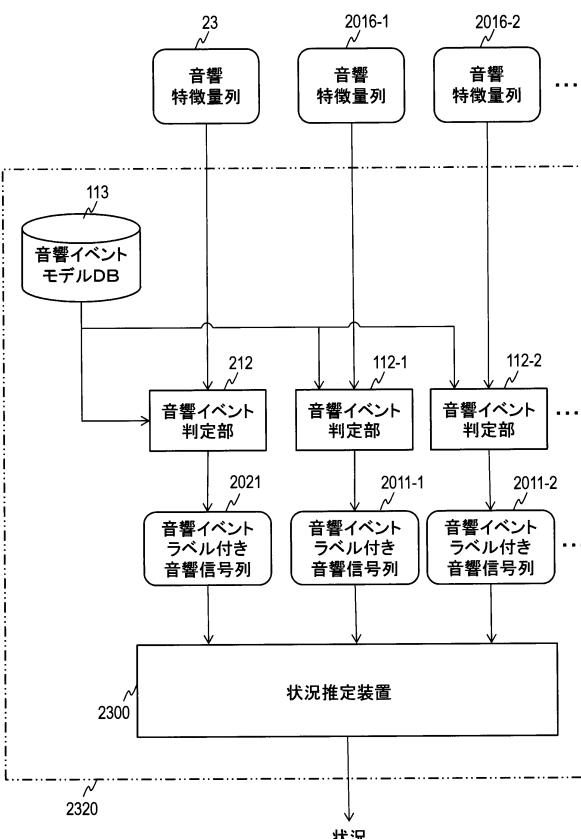


図31

図32

フロントページの続き

特許法第30条第2項適用 発行日 平成25年3月5日 刊行物名 日本音響学会 2013年春季研究発表会「講演論文集」 講演要旨・講演論文CD-ROM 発表題目 「音響イベント列の確率的生成モデルを利用した音響トピックとユーザ行動との関係分析」(1-P-3 P.825~826) 開催日 平成25年3月13日~平成25年3月15日 集会名 「日本音響学会 2013年春季研究発表会」 開催場所 東京工科大学八王子キャンパス 公開日 平成25年3月13日 公開者 井本桂右・島内未廣・植松尚・大室仲・羽田陽一

審査官 安田 勇太

(56)参考文献 特開2014-048521(JP,A)

特開2014-048522(JP,A)

特開2014-048523(JP,A)

特開2006-279898(JP,A)

井本桂右, 音響イベントを用いた人の行動の確率的生成モデルによる行動識別手法の検討, 日本音響学会 2012年 秋季研究発表会講演論文集, 2012年 9月21日, pp.975-976

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/34

G10L 25/00 - 25/93

G06F 19/00

G06N 5/04