

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7619465号
(P7619465)

(45)発行日 令和7年1月22日(2025.1.22)

(24)登録日 令和7年1月14日(2025.1.14)

(51)国際特許分類

F I

A 6 1 B 3/11 (2006.01)

A 6 1 B 3/11

請求項の数 10 (全19頁)

(21)出願番号	特願2023-539458(P2023-539458)	(73)特許権者	000004226
(86)(22)出願日	令和3年8月4日(2021.8.4)		日本電信電話株式会社
(86)国際出願番号	PCT/JP2021/028988		東京都千代田区大手町一丁目5番1号
(87)国際公開番号	WO2023/012941	(74)代理人	100121706
(87)国際公開日	令和5年2月9日(2023.2.9)		弁理士 中尾 直樹
審査請求日	令和6年1月25日(2024.1.25)	(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	鈴木 雄太
			東京都千代田区大手町一丁目5番1号
			日本電信電話株式会社内
		(72)発明者	リャオ シンイ
			東京都千代田区大手町一丁目5番1号
			日本電信電話株式会社内
			最終頁に続く

(54)【発明の名称】 聴覚注意状態推定装置、学習装置、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】

互いに異なる複数の音源の位置から発せられる互いに異なる複数の視覚刺激パターンのそれぞれと利用者の瞳孔径変化量との相関の強さに基づく特徴量を得る特徴量抽出部と、
前記特徴量を用い、前記音源からの音について前記利用者が聴覚的に注意を向けた先を推定する推定部と、
を有する聴覚注意状態推定装置。

【請求項2】

請求項1の聴覚注意状態推定装置であって、
前記複数の視覚刺激パターンのそれぞれは、周期的に時間変化する視覚刺激のパターンであり、
前記特徴量は、少なくとも、複数の周波数領域視覚刺激信号それぞれのピーク周波数および／または前記ピーク周波数の近傍における周波数領域瞳孔径変化量信号の大きさを表し、
前記複数の周波数領域視覚刺激信号のそれぞれは、前記複数の視覚刺激パターンのそれぞれを表す時系列信号を周波数領域に変換して得られる信号であり、
前記周波数領域瞳孔径変化量信号は、前記瞳孔径変化量を表す時系列信号を周波数領域に変換して得られる信号である、聴覚注意状態推定装置。

【請求項3】

請求項1の聴覚注意状態推定装置であって、

前記複数の視覚刺激パターンのそれぞれは、時間変化する視覚刺激のパターンであり、
前記特徴量は、少なくとも、前記複数の視覚刺激パターンのそれぞれを表す時系列信号
に対応する系列と前記瞳孔径変化量を表す時系列信号に対応する系列とに対する相互相関
関数の各最大値を表す、聴覚注意状態推定装置。

【請求項 4】

請求項 1 から 3 の何れかの聴覚注意状態推定装置であって、

前記複数の音源のうち、前記利用者の瞳孔径変化量との相関が強い視覚刺激パターンに
対応する音源ほど、前記推定部が、前記利用者が聴覚的に注意を向けた先であると推定す
る頻度が高い、聴覚注意状態推定装置。

【請求項 5】

請求項 1 から 3 の何れかの聴覚注意状態推定装置であって、

前記複数の音源は、第 1 音源を含み、

前記複数の視覚刺激パターンは、前記第 1 音源に対応する第 1 視覚刺激パターンを含み、

前記第 1 視覚刺激パターンと前記利用者の瞳孔径変化量との相関の強さが所定値以上で
ある場合に、前記推定部が、前記利用者が聴覚的に注意を向けた先が前記第 1 音源または
前記第 1 音源の近傍であると推定する、聴覚注意状態推定装置。

【請求項 6】

請求項 1 から 3 の何れかの聴覚注意状態推定装置であって、

前記複数の音源は、第 1 音源および前記第 1 音源以外の第 2 音源を含み、

前記複数の視覚刺激パターンは、前記第 1 音源に対応する第 1 視覚刺激パターンおよび
前記第 2 音源に対応する第 2 視覚刺激パターンを含み、

前記第 1 視覚刺激パターンと前記利用者の瞳孔径変化量との相関の強さが、前記第 2 視
覚刺激パターンと前記利用者の瞳孔径変化量との相関の強さよりも強い場合に、前記推定
部が、前記利用者が聴覚的に注意を向けた先が前記第 1 音源または前記第 1 音源の近傍で
あると推定する、聴覚注意状態推定装置。

【請求項 7】

互いに異なる複数の学習用音源の位置から発せられる互いに異なる複数の学習用視覚刺
激パターンのそれぞれと学習用瞳孔径変化量との相関の強さに基づく学習用特徴量と、前
記学習用音源からの学習用音について聴覚的に注意が向けられた先を表す正解情報と、を
対応付けた学習データを用いた学習処理により、

互いに異なる複数の音源の位置から発せられる互いに異なる複数の視覚刺激パターンのそ
れぞれと利用者の瞳孔径変化量との相関の強さに基づく特徴量を入力とし、前記音源から
の音について前記利用者が聴覚的に注意を向けた先を推定する推定モデルを得る学習部を
有する学習装置。

【請求項 8】

聴覚注意状態推定装置による聴覚注意状態推定方法であって、

特徴量抽出部が、互いに異なる複数の音源の位置から発せられる互いに異なる複数の視
覚刺激パターンのそれぞれと利用者の瞳孔径変化量との相関の強さに基づく特徴量を得る
特徴量抽出ステップと、

推定部が、前記特徴量を用い、前記音源からの音について前記利用者が聴覚的に注意を
向けた先を推定する推定ステップと、
を有する聴覚注意状態推定方法。

【請求項 9】

学習装置による学習方法であって、

学習部が、互いに異なる複数の学習用音源の位置から発せられる互いに異なる複数の学
習用視覚刺激パターンのそれぞれと学習用瞳孔径変化量との相関の強さに基づく学習用特
徴量と、前記学習用音源からの学習用音について聴覚的に注意が向けられた先を表す正解
情報と、を対応付けた学習データを用いた学習処理により、

互いに異なる複数の音源の位置から発せられる互いに異なる複数の視覚刺激パターンのそ
れぞれと利用者の瞳孔径変化量との相関の強さに基づく特徴量を入力とし、前記音源から

10

20

30

40

50

の音について前記利用者が聴覚的に注意を向けた先を推定する推定モデルを得る学習ステップを有する学習方法。

【請求項 10】

請求項 1 から 6 の何れかの聴覚注意状態推定装置または請求項 7 の学習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、聴覚的な注意状態を推定する技術に関する。

【背景技術】

10

【0002】

ヒトは、視覚・聴覚・触覚などといった感覚に対する特定の入力刺激について選択的に注意を向けることで、認知すべき情報の取捨選択を意識または無意識下で行っている。非特許文献 1 では、光源の ON/OFF によって誘発される瞳孔径の縮小・散大（瞳孔振動、または PFT (Pupillary Frequency Tagging)）を利用し、利用者が視覚的な注意を向けた先を推定している。

【先行技術文献】

【非特許文献】

【0003】

【文献】Naber, M., Alvarez, G. A., and Nakayama, K., "Tracking the allocation of attention using human pupillary oscillations," [online], 2013年12月10日, Frontiers in Psychology, 4., [2021年7月15日検索], インターネット<<http://doi.org/10.3389/fpsyg.2013.00919>>

20

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかし、利用者が聴覚的な注意を向けた先と瞳孔径の変化との関係は知られておらず、瞳孔径の変化に基づいて、利用者が聴覚的な注意を向けた先を推定する方法も知られていない。

【0005】

30

本発明は、瞳孔径の変化に基づいて、利用者が聴覚的な注意を向けた先を推定する手法を提供する。

【課題を解決するための手段】

【0006】

互いに異なる複数の音源に対応する互いに異なる複数の視覚刺激パターンのそれぞれと利用者の瞳孔径変化量との相関の強さに基づく特徴量を得、当該特徴量を用い、音源からの音について利用者が聴覚的に注意を向けた先を推定する。

【発明の効果】

【0007】

本発明では、瞳孔径の変化に基づいて、利用者が聴覚的な注意を向けた先を推定できる。

40

【図面の簡単な説明】

【0008】

【図 1】図 1 は、実施形態の聴覚注意状態推定システムの機能構成を例示したブロック図である。

【図 2】図 2 は、実施形態の学習装置の機構構成を例示したブロック図である。

【図 3】図 3 は、実施形態の聴覚注意状態推定装置の機構構成を例示したブロック図である。

【図 4】図 4 は、実験内容を説明するための概念図である。

【図 5】図 5 は、実験結果を例示したグラフである。

【図 6】図 6 は、実施形態の学習装置および聴覚注意状態推定装置のハードウェア構成を

50

例示したブロック図である。

【発明を実施するための形態】

【0009】

以下、図面を参照して本発明の実施の形態を説明する。

[原理]

まず、原理を説明する。本発明は、ヒトが、音源に対応する視覚刺激を注視することなく、当該音源からの音について聴覚的に注意を向けているだけでも瞳孔反応を示すという新しい自然法則（生理学上の法則）の発見に基づく。まず、この発見に至った実験結果を示す。

【0010】

<実験内容>

図4に例示するように、この実験では4個の音源（この実験ではスピーカー）140-1, ..., 140-4を配置し、各音源140-i（ただしi=1, ..., 4）に光源130-i（この実験ではLED光源）を設置した。光源130-1, ..., 130-4は、互いに異なる周波数（明滅周波数）で明滅（ON/OFF）し、これによって互いに異なる視覚刺激パターンの視覚刺激を提示する。この実験では、光源130-1, 130-2, 130-3, 130-4の明滅周波数を、それぞれ1.50Hz, 1.75Hz, 2.00Hz, 2.25Hzとした。また4個の音源140-1, ..., 140-4からは互いに異なる音が同時に発せられる。被検者100は、何れかの音源140-iから発せられた音に聴覚的に注意を向ける課題（タスク）を実行した。例えば、音源140-1, ..., 140-4から、複数のカテゴリに基づいた単語群を含む音声ランダムに発せられ、被検者100は何れか1つの音源140-iから発せられる音声の中の指定されたカテゴリの単語数を数えるという課題を実行した。例えば、4個の音源140-1, ..., 140-4から、以下の音声と同時に出力され、被検者100は音源140-1から発せられた音声に含まれた数字（3,8,9,6）または色（白色, 桃色, 赤色, 黒色）を答えるという課題を実行した。この場合、被検者100は音源140-1から発せられた音に聴覚的な注意を向けることになる。

音源140-1: 「それではクマは白色の3へどうぞ」「それではトラは桃色の8へどうぞ」「それではネコは赤色の9へどうぞ」「それではイヌは黒色の6へどうぞ」

音源140-2: 「それではシカは虹色の2へどうぞ」を4回

音源140-3: 「それではトラは青色の9へどうぞ」を4回

音源140-4: 「それではブタは桃色の7へどうぞ」を4回

被検者100は、各音源140-1, 2, 3, 4から発せられた音に聴覚的な注意を向ける課題を順番に実行し、課題を実行している被検者100の瞳孔径が瞳孔径取得装置150（この実験ではアイトラッカー）で計測された。課題の実行と瞳孔径の計測とは複数人の被検者100について複数回行われ、各被検者100が各音源140-iから発せられた音に聴覚的な注意を向けているときの瞳孔径が測定された。

【0011】

<実験結果>

図5に、上述した課題を実行している被検者100の瞳孔径変化量を表す時系列信号を周波数領域に変換して得られた信号（周波数領域瞳孔径変化量信号）を例示する。図5の横軸は周波数（Hz）を表し、縦軸は周波数領域瞳孔径変化量信号のパワー[a.u.]を表す。図5では、被検者が視覚的に注意を向けた音源140-i（ただしi=1, ..., 4）の違い（すなわち、音源140-iに対応する光源130-iの明滅周波数の違い）を線の種別によって区別している。太い線はパワーの平均値を表し、細い線はパワーの分布を表す。

【0012】

図5に例示するように、被検者100音源140-iから発せられる音に聴覚的な注意を向けた場合、当該被検者100の瞳孔径変化量に基づく周波数領域瞳孔径変化量信号のパワーのピークは、音源140-iに設けられた光源130-iの明滅周波数またはその近傍となる。例えば、被検者100が明滅周波数1.50[Hz]で明滅する光源130-1が設けられた音源140-1から発せられる音に聴覚的な注意を向けた場合、当該被検者100

10

20

30

40

50

0 の瞳孔径変化量に基づく周波数領域瞳孔径変化量信号のパワーのピークp1は1.50[Hz]またはその近傍に現れる。被検者100が明滅周波数1.75[Hz]で明滅する光源130-2が設けられた音源140-2から発せられる音に聴覚的な注意を向けた場合、当該被検者100の瞳孔径変化量に基づく周波数領域瞳孔径変化量信号のパワーのピークp2は1.75[Hz]またはその近傍に現れる。被検者100が明滅周波数2.00[Hz]で明滅する光源130-3が設けられた音源140-3から発せられる音に聴覚的な注意を向けた場合、当該被検者100の瞳孔径変化量に基づく周波数領域瞳孔径変化量信号のパワーのピークp3は2.00[Hz]またはその近傍に現れる。被検者100が明滅周波数2.25[Hz]で明滅する光源130-4が設けられた音源140-4から発せられる音に聴覚的な注意を向けた場合、当該被検者100の瞳孔径変化量に基づく周波数領域瞳孔径変化量信号のパワーのピークp4は2.25[Hz]またはその近傍に現れる。すなわち、被検者100が音源140-iから発せられた音に聴覚的な注意を向けたときの瞳孔径変化量と、その音源140-iに対応する光源130-1から発せられた視覚刺激パターンとの間には相関がある。ここで、課題の実行中、被検者100は光源130-1を注視するよう指示されていない。よって、この相関は、被検者100が光源130-1を注視していなくても成り立つと予想される。また、音源に対応する視覚刺激パターンは、時間変化する視覚刺激のパターンであり、例えば、周期的に時間変化する視覚刺激のパターンであってもよいし、非周期的に時間変化する視覚刺激のパターンであってもよい。視覚刺激パターンは時間変化する輝度（明るさ）のパターンであってもよいし、時間変化する色彩のパターンであってもよいし、時間変化する模様のパターンであってもよいし、時間変化する形状のパターンであってもよい。音源に対応する視覚刺激パターンは光源によって作られるものに限られず、ディスプレイや投影機によって表示や投影されるものであってもよいし、音源自体や音源周囲の環境（例えば、輝度変化）が作り出すものであってもよい。

10

20

【0013】

以上のように、ヒトは、音源から発せられた音に聴覚的に注意を向けただけでも、音源に対応する視覚刺激パターンに応じた瞳孔反応を示す。各実施形態では、この自然法則を利用し、互いに異なる複数の音源に対応する互いに異なる複数の視覚刺激パターンのそれぞれと利用者の瞳孔径変化量との相関の強さに基づく特徴量を得、当該特徴量を用い、音源からの音について利用者が聴覚的に注意を向けた先を推定する。

【0014】

[第1実施形態]

次に、第1実施形態を説明する。

<全体構成>

図1に例示するように、本実施形態の聴覚注意状態推定システム1は、学習装置11、聴覚注意状態推定装置12、複数の視覚刺激発生装置13-1, ..., 13-N、複数の音源装置14-1, ..., 14-N、および瞳孔径取得装置15を有し、利用者10の瞳孔径変化量に基づいて、当該利用者10が聴覚的に注意を向けた先を推定する。ただし、Nは1以上の整数であり、例えば、Nは2以上の整数である。

【0015】

<学習装置11>

図2に例示するように、本実施形態の学習装置11は、入力部111、記憶部112、学習部113、出力部114、および制御部117を有する。以降、説明を省略するが、学習装置11は、制御部117の制御に基づいて各処理を実行し、学習装置11への入力データおよび各処理で得られたデータは記憶部112に格納され、必要に応じて読み出されて使用される。

【0016】

<聴覚注意状態推定装置12>

図3に例示するように、本実施形態の聴覚注意状態推定装置12は、入力部121、記憶部122、視覚刺激制御部123、聴覚情報制御部124、特徴量抽出部125、推定部126、および制御部127を有する。以降、説明を省略するが、聴覚注意状態推定装

30

40

50

置 1 2 は、制御部 1 2 7 の制御に基づいて各処理を実行し、聴覚注意状態推定装置 1 2 への入力データおよび各処理で得られたデータは記憶部 1 2 2 に格納され、必要に応じて読み出されて使用される。

【 0 0 1 7 】

< 音源装置 1 4 - n >

本実施形態の音源装置 1 4 - n (ただし、 $n = 1, \dots, N$) は、音 SO(Info-n)を発する音源である。音源装置 1 4 - n の例は、スピーカなどであるがこれは本発明を限定しない。空間的に所望の位置に音源を配置可能な装置であれば、どのような装置を音源装置 1 4 - n として用いてもよい。音源装置 1 4 - 1, ..., 1 4 - N は互いに異なり、音源装置 1 4 - 1, ..., 1 4 - N は互いに異なる位置に音源を配置する。例えば、利用者 1 0 からの音源装置 1 4 - 1, ..., 1 4 - N の方向または音源装置 1 4 - 1, ..., 1 4 - N によって配置される音源の方向は互いに異なる。音源装置 1 4 - 1, ..., 1 4 - N は同時に音 SO(Info-1), ..., SO(Info-N)を発してもよいし、音源装置 1 4 - 1, ..., 1 4 - N のうちの一部の音源装置が他の音源装置と異なるタイミングで音 SO(Info-n)を発してもよい。ただし、異なる音源装置から同時に発せられる音は互いに異なることが好ましい。音源装置 1 4 - 1, ..., 1 4 - N から発せられる音 SO(Info-1), ..., SO(Info-N)は、音声、音楽、環境音、呼び出し音、アラーム音等どのようなものであってもよい。

10

【 0 0 1 8 】

< 視覚刺激発生装置 1 3 - n >

本実施形態の視覚刺激発生装置 1 3 - n (ただし、 $n = 1, \dots, N$) は、音源装置 1 4 - n に対応する視覚刺激パターン VS(Sig-n)を提示 (表示) する装置である。利用者 1 0 が音源装置 1 4 - n と視覚刺激発生装置 1 3 - n との対応関係を知覚できるのであれば、どのように視覚刺激発生装置 1 3 - n が配置・構成されていてもよい。例えば、視覚刺激発生装置 1 3 - n が、音源装置 1 4 - n の近傍に配置されてもよいし、音源装置 1 4 - n に接触して配置されてもよいし、音源装置 1 4 - n に固定されてもよいし、音源装置 1 4 - n と一体に構成されてもよい。本実施形態の視覚刺激パターン VS(Sig-n)のそれぞれは、周期的に時間変化する視覚刺激のパターンである。例えば、実施形態の視覚刺激パターン VS(Sig-n)は、周期的に時間変化する輝度 (明るさ) のパターンであってもよいし、周期的に明滅 (ON/OFF) を繰り返すパターンであってもよいし、周期的に時間変化する色彩のパターンであってもよいし、周期的に時間変化する模様のパターンであってもよいし、周期的に時間変化する形状のパターンであってもよい。すなわち、視覚刺激発生装置 1 3 - n は、周期的に時間変化する輝度 (明るさ) を提示してもよいし、周期的に明滅を繰り返す光を提示してもよいし、周期的に時間変化する色彩を提示してもよいし、周期的に時間変化する模様を提示してもよいし、周期的に時間変化する形状を提示してもよい。このような視覚刺激パターン VS(Sig-n)を視覚的に提示できるのであれば、視覚刺激発生装置 1 3 - n はどのようなものであってもよい。例えば、視覚刺激発生装置 1 3 - n は、LED 光源であってもよいし、レーザ光発生器であってもよいし、ディスプレイであってもよいし、投影機であってもよい。また N 個の視覚刺激発生装置 1 3 - 1, ..., 1 3 - N から提示される視覚刺激パターン VS(Sig-1), ..., VS(Sig-N)は互いに異なる。本実施形態の場合、視覚刺激発生装置 1 3 - 1, ..., 1 3 - N から提示される視覚刺激パターン VS(Sig-1), ..., VS(Sig-N)は互いに異なる周波数分布を持つ。例えば、本実施形態の視覚刺激発生装置 1 3 - 1, ..., 1 3 - N から提示される視覚刺激パターン VS(Sig-1), ..., VS(Sig-N)を表す時系列信号 (例えば、輝度の時系列信号、色彩の時系列信号、模様の時系列信号、形状の時系列信号など) を周波数領域に変換 (例えば、フーリエ変換) して得られる信号 (周波数領域視覚刺激信号) のピークの周波数 (ピーク周波数) は、互いに異なる。

20

30

40

【 0 0 1 9 】

< 瞳孔径取得装置 1 5 >

本実施形態の瞳孔径取得装置 1 5 は、利用者 1 0 の瞳孔径 Pub を測定する装置である。例えば、瞳孔径取得装置 1 5 は利用者 1 0 の目の動きを撮影するカメラと、当該カメラで

50

撮影された映像から利用者 10 の瞳孔径Pubを取得して出力する装置である。瞳孔径取得装置 15 の一例は、市販のアイトラッカー等である。

【0020】

<処理の全体>

図 1 に例示するように、学習装置 11 には学習データTが入力され、学習装置 11 は学習データTを用いた学習処理によって推定モデル $M(\theta)$ を得、当該推定モデル $M(\theta)$ を特定するモデルパラメータ θ を出力する。モデルパラメータ θ は聴覚注意状態推定装置 12 に入力される。聴覚注意状態推定装置 12 は、出力情報Info-1,...,Info-Nを音源装置 14-1,...,14-Nにそれぞれ出力し、音源装置 14-1,...,14-Nは出力情報Info-1,...,Info-Nに基づく音SO(Info-1),...,SO(Info-N)をそれぞれ出力する。また聴覚注意状態推定装置 12 は、出力情報Sig-1,...,Sig-Nを視覚刺激発生装置 13-1,...,13-Nにそれぞれ出力し、視覚刺激発生装置 13-1,...,13-Nは、出力情報Sig-1,...,Sig-Nに基づく視覚刺激パターンVS(Sig-1),...,VS(Sig-N)をそれぞれ提示する。利用者 10 は、少なくとも何れかの音源装置 14-nから発せられた音SO(Info-n)に対して聴覚的に注意を向け、瞳孔径取得装置 15 は利用者 10 の瞳孔径Pubを取得して聴覚注意状態推定装置 12 に送る。聴覚注意状態推定装置 12 は、出力情報Sig-1,...,Sig-Nに基づいて提示された視覚刺激パターンVS(Sig-1),...,VS(Sig-N)のそれぞれと、瞳孔径Pubに基づいて得られる利用者 10 の瞳孔径変化量と、の相関の強さに基づく特徴量fを得る。さらに、聴覚注意状態推定装置 12 は、特徴量fを用い、モデルパラメータ θ によって特定される推定モデル推定モデル $M(\theta)$ (機械学習モデル)に基づき、音源からの音について利用者 10 が聴覚的に注意を向けた先の推定結果 $E=M(\theta;f)$ を得て出力する。以下、詳細に説明する。

【0021】

<学習装置 11 の処理>

学習装置 11 (図 2) の入力部 111 には、学習データTが入力され、学習データTは記憶部 112 に格納される。学習データTは、互いに異なるN個(複数)の学習用音源に対応する互いに異なるN個(複数)の学習用視覚刺激パターンTVS(Sig-1),...,TVS(Sig-N)のそれぞれと学習用瞳孔径変化量TVPubとの相関の強さに基づく学習用特徴量Tf_jと、学習用音源からの音(学習用音)について聴覚的に注意が向けられた先を表す正解情報Ta_jと、を対応付けたデータ(Tf₁, Ta₁),..., (Tf_J, Ta_J)である。ここで、Jは正の整数であり、j=1,...,Jである。

【0022】

学習用音は、音声、音楽、環境音、呼び出し音、アラーム音等どのような音であってもよい。学習用音の具体例は、前述した音源装置 14-nから発せられる音と同一である。

【0023】

N個の学習用音源の具体例は、前述した音源装置 14-1,...,14-Nと同一である。しかし、N個の学習用音源は学習用音を発するN個の装置であればよく、前述した音源装置 14-1,...,14-Nと異なる装置がN個の学習用音源であってもよい。

【0024】

N個の学習用視覚刺激パターンTVS(Sig-1),...,TVS(Sig-N)は、前述した視覚刺激パターンVS(Sig-1),...,VS(Sig-N)とそれぞれ同一であり、周期的に時間変化する視覚刺激のパターンである。例えば、N個の学習用視覚刺激パターンTVS(Sig-1),...,TVS(Sig-N)は、前述した視覚刺激発生装置 13-1,...,13-Nから提示される視覚刺激のパターンとそれぞれ同一であり、学習用視覚刺激パターンTVS(Sig-1),...,TVS(Sig-N)の具体例は、前述した視覚刺激パターンVS(Sig-1),...,VS(Sig-N)の具体例とそれぞれ同一である。

【0025】

学習用瞳孔径変化量TVPubは、学習用音および学習用視覚刺激パターンTVS(Sig-1),...,TVS(Sig-N)が提示されている学習用利用者が、何れかの学習用音TSO_n(ただし、n{1,...,N})について聴覚的に注意を向けているときにおける、当該学習用利用者の瞳孔径変化量である。学習用利用者が何れかの学習用音TSO_nについて聴覚的に注意を向けてい

10

20

30

40

50

るときとは、例えば、学習用利用者が前述した課題（タスク）を実行しているときである。学習用瞳孔径変化量 TV_{Pub} は、例えば、瞳孔径取得装置15によって計測された学習用利用者の瞳孔径 TP_{Pub} に基づいて得られるが、その他の装置で計測された学習用利用者の瞳孔径 TP_{Pub} に基づいて得られてもよい。以下に瞳孔径 TP_{Pub} から学習用瞳孔径変化量を得る方法を例示する。

(1.1) まず、瞳孔径 TP_{Pub} の時系列データに対する前処理が行われ、前処理後の瞳孔径 $TP_{Pub'}$ の時系列データが得られる。前処理としては、例えば、線形補間や2次スプライン補間などを用い、瞳孔径 TP_{Pub} の時系列データから学習用利用者の瞬き等によって欠損した部分を補間する処理を例示できる。さらに、学習用視覚刺激パターン $TVS(Sig-1), \dots, TVS(Sig-N)$ の明滅周波数に応じたローパスフィルタ（例えば、すべての明滅周波数を含む帯域を通すローパスフィルタ）を補完後の瞳孔径 TP_{Pub} の時系列データに適用し、ノイズの軽減を行ってもよい。

10

(1.2) 次に、学習用利用者が何れかの学習用音 T_{SO_n} について聴覚的に注意を向けているときの瞳孔径 $TP_{Pub'}$ の時系列データから、聴覚的に注意を向ける前の瞳孔径 $TP_{Pub'}$ の平均値が減算され、 z 値による標準化が行われて学習用瞳孔径変化量 TV_{Pub} の時系列データが得られる。

【0026】

学習用特徴量 Tf_j は、 N 個の学習用視覚刺激パターン $TVS(Sig-1), \dots, TVS(Sig-N)$ のそれぞれと学習用瞳孔径変化量 TV_{Pub} との相関の強さに基づくものであれば、どのようなものであってもよい。以下に学習用特徴量 Tf_j を例示する。

20

【0027】

(2.1) 複数の学習用周波数領域視覚刺激信号 $TFVS(Sig-1), \dots, TFVS(Sig-N)$ それぞれのピーク周波数および／またはピーク周波数の近傍における学習用周波数領域瞳孔径変化量信号 TFV_{Pub} の大きさ（例えば、 TFV_{Pub} の大きさのピーク値）を表す学習用特徴量 Tf_j が用いられてもよい。ここで、複数の学習用周波数領域視覚刺激信号 $TFVS(Sig-1), \dots, TFVS(Sig-N)$ のそれぞれは、複数の学習用視覚刺激パターン $TVS(Sig-1), \dots, TVS(Sig-N)$ のそれぞれを表す時系列信号を周波数領域に変換（例えば、フーリエ変換）して得られる信号である。また、学習用周波数領域瞳孔径変化量信号 TFV_{Pub} は、学習用瞳孔径変化量 TV_{Pub} を表す時系列信号を周波数領域に変換して得られる信号である。また、「 α の大きさ」とは、 α の振幅の絶対値であってもよいし、 α のパワー（ α の振幅の2乗）であってもよいし、その他、 α の絶対値に対して単調増加する値であってもよい。例えば、図5の例の場合、ピーク $p1, \dots, p4$ またはそれらの近傍での各パワー（各パワーのピーク値）を要素に含む学習用特徴量 Tf_j が用いられてもよい。

30

【0028】

ここで、ピーク周波数および／またはピーク周波数の近傍における学習用周波数領域瞳孔径変化量信号 TFV_{Pub} の大きさ（例えば、 TFV_{Pub} の大きさのピーク値）が大きいほど、当該ピーク周波数に対応する学習用視覚刺激パターン $TVS(Sig-n)$ と学習用瞳孔径変化量 TV_{Pub} との相関が強い。

【0029】

(2.2) 上述したピーク周波数および／またはピーク周波数の近傍における学習用周波数領域瞳孔径変化量信号 TFV_{Pub} の大きさに加え、または、これに代えて、その他の情報が学習用特徴量 Tf_j に含まれてもよい。例えば、(1)複数の学習用周波数領域視覚刺激信号 $TFVS(Sig-1), \dots, TFVS(Sig-N)$ それぞれのピーク周波数の倍周波数またはそれらの近傍における学習用周波数領域瞳孔径変化量信号 TFV_{Pub} の大きさ（例えば、大きさのピーク値）、(2) N 個の学習用視覚刺激パターン $TVS(Sig-1), \dots, TVS(Sig-N)$ のそれぞれの位相変化と学習用瞳孔径変化量 TV_{Pub} の位相変化の同期度、(3) N 個の学習用視覚刺激パターン $TVS(Sig-1), \dots, TVS(Sig-N)$ のそれぞれを表す時系列信号に対応する系列 $TSS(Sig-1), \dots, TSS(Sig-N)$ のそれぞれと学習用瞳孔径変化量 TV_{Pub} を表す時系列信号に対応する系列 TPS とに対する相互相関関数の各最大値 $CCF_{max}(TSS(Sig-1), TPS), \dots, CCF_{max}(TSS(Sig-N), TPS)$ などが、学習用特徴量 Tf_j に含まれてもよい。ここで、「時系列信号に対応する

40

50

系列」は、例えば、時系列信号そのものであってもよいし、時系列信号を周波数領域に変換して得られる系列であってよいし、その他の時系列信号の関数値の系列であってよい。また、系列 $\alpha 1$ と系列 $\alpha 2$ との相互相関関数の最大値とは、可変遅延量 τ についての系列 $\alpha 1$ と系列 $\alpha 2$ との相互相関関数値のうち、最大の値を意味する。

【0030】

ここで、ピーク周波数の倍周波数および/またはその近傍における学習用周波数領域瞳孔径変化量信号TFVPubの大きさ(例えば、TFVPubの大きさのピーク値)が大きいほど、当該ピーク周波数に対応する学習用視覚刺激パターンTVS(Sig-n)と学習用瞳孔径変化量TVPubとの相関が強い。また、学習用瞳孔径変化量TVPubとの同期度が高い位相変化を持つ学習用視覚刺激パターンTVS(Sig-n)ほど、学習用瞳孔径変化量TVPubとの相関が強い。また、相互相関関数の最大値CCFmax(TSS(Sig-n), TPS)が大きい学習用視覚刺激パターンTVS(Sig-n)ほど、学習用瞳孔径変化量TVPubとの相関が強い。

10

【0031】

前述のように、学習用瞳孔径変化量TVPubは、学習用利用者が学習用音TSO_nに聴覚的に注意を向けているときにおける当該学習用利用者の瞳孔径変化量であり、正解情報Ta_jは当該学習用音TSO_nを表す情報である。すなわち、正解情報Ta_jは、学習用瞳孔径変化量TVPubに対応する学習用音TSO_nを表す。例えば、正解情報Ta_jは、音源(例えば、学習用音TSO_nを発する装置)を表す情報(例えば、インデックス)であってよいし、学習用音TSO_nに対応する学習用視覚刺激パターンTVS(Sig-n)を表す情報(例えば、明滅周波数)であってよいし、学習用音TSO_nを表す情報であってよい。上述のように、学習用特徴量Tf_jは学習用瞳孔径変化量TVPubに対応するため、正解情報Ta_jは学習用特徴量Tf_jと対応付けられる(ステップS111)。

20

【0032】

学習部113は、記憶部112から読み込んだ学習データTを用いた学習処理(機械学習)によって推定モデルM(θ)を得、当該推定モデルM(θ)を特定するモデルパラメータ θ を出力する。推定モデルM(θ)は、互いに異なるN個(複数)の音源に対応する互いに異なるN個(複数)の視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)のそれぞれと利用者の瞳孔径変化量VPubとの相関の強さに基づく特徴量fを入力とし、音源からの音について利用者が聴覚的に注意を向けた先(視覚的な注意方向)を推定するためのモデルである。特徴量fの構成は、学習用音が音に置換され、学習用音源が音源に置換され、学習用視覚刺激パターンTVS(Sig-1), ..., TVS(Sig-N)が視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)に置換され、学習用利用者が利用者に置換され、学習用利用者の学習用瞳孔径変化量TVPubが利用者の瞳孔径変化量VPubに置換される以外、上述した学習用特徴量Tf_jの構成と同一である。

30

【0033】

推定モデルM(θ)は、瞳孔径変化量VPubと相関が高くなる視覚刺激パターンVS(Sig-n)に対応する音源から発せられ音を、利用者が聴覚的に注意を向けた先と推定するよう構成される。このような推定モデルM(θ)を用いて推定される「利用者が聴覚的に注意を向けた先」は、例えば、以下の(3.1), (3.2), (3.3)の少なくとも何れかのようになる。

【0034】

(3.1)瞳孔径変化量VPubとの相関が高い視覚刺激パターンVS(Sig-n)に対応する音源から発せられ音ほど、利用者が聴覚的に注意を向けた先と推定される頻度(確率)が高くなる。

40

【0035】

(3.2)視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)に含まれる視覚刺激パターンVS(Sig-n)(第1視覚刺激パターン)が音源AS₁(第1音源)に対応し、視覚刺激パターンVS(Sig-n)と利用者の瞳孔径変化量VPubとの相関の強さが所定値以上である場合に、利用者が聴覚的に注意を向けた先が音源AS₁または音源AS₁の近傍であると推定される。

【0036】

(3.3)対象となるN個(複数)の音源が音源AS₁(第1音源)および音源AS₂(第2音源)を含み、視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)に含まれる視覚刺激パターンVS(Si

50

$g-n_1$) (第1視覚刺激パターン) が音源 AS_1 に対応し、視覚刺激パターン $VS(Sig-n_2)$ (第2視覚刺激パターン) が音源 AS_2 に対応し、視覚刺激パターン $VS(Sig-n_1)$ と瞳孔径変化量 V_{Pub} との相関の強さが、視覚刺激パターン $VS(Sig-n_2)$ と瞳孔径変化量 V_{Pub} との相関の強さよりも強い場合に、利用者が聴覚的に注意を向けた先が音源 AS_1 または音源 AS_1 の近傍であると推定される。ここで、 $n_1, n_2 \in \{1, \dots, N\}$ である。

【0037】

なお、推定モデル $M(\theta)$ を用いて推定される「利用者が聴覚的に注意を向けた先」は、音源(例えば、音を発する装置)を表す情報(例えば、インデックス)であってもよいし、音 SO_n に対応する視覚刺激パターン $VS(Sig-n)$ を表す情報(例えば、明滅周波数)であってもよいし、音 SO_n を表す情報であってもよいし、これらの方向を表す情報であってもよい。また、推定モデル $M(\theta)$ によって、1個の「利用者が聴覚的に注意を向けた先」が推定されてもよいし、複数の「利用者が聴覚的に注意を向けた先」が推定されてもよいし、「利用者が聴覚的に注意を向けた先」の確率が推定されてもよい。

10

【0038】

推定モデル $M(\theta)$ は、どのような方式に基づくものであってもよい。例えば、K-NN(k-nearest neighbor algorithm), SVM(support vector machine), ディープラーニング, 隠れマルコフモデルなどに基づく推定モデル $M(\theta)$ を例示できる。学習処理の具体的な方法は、推定モデル $M(\theta)$ の方式に応じた公知な方法を用いればよい。一般的には、最初に適当な暫定モデルパラメータ θ' の初期値を設定し、その後、推定モデル $M(\theta')$ に学習用特徴量 Tf_j を適用して得られる結果と正解情報 Taj との誤差が小さくなるように、暫定モデルパラメータ θ' を更新する処理を繰り返し、所定の終了条件を満たした時点の暫定モデルパラメータ θ' をモデルパラメータ θ とする。

20

【0039】

学習部113から出力されたモデルパラメータ θ は出力部114に送られ、出力部114はモデルパラメータ θ を聴覚注意状態推定装置12に送る(ステップS113)。

【0040】

<聴覚注意状態推定装置12の処理>

学習装置11から送られたモデルパラメータ θ は聴覚注意状態推定装置12(図3)の入力部121に入力され、記憶部122に格納される。また、記憶部122には、視覚刺激パターン $VS(Sig-1), \dots, VS(Sig-N)$ を表す出力情報 $Sig-1, \dots, Sig-N$ が格納され、音 $SO(Info-1), \dots, SO(Info-N)$ を表す出力情報 $Info-1, \dots, Info-N$ が格納される(ステップS122)。

30

【0041】

聴覚情報制御部124は、記憶部122から出力情報 $Info-1, \dots, Info-N$ を読み込んで音源装置14-1, $\dots$, 14-Nにそれぞれ送る。各音源装置14-nは、送られた出力情報 $Info-n$ に基づく音 $SO(Info-n)$ を提示(出力)する(ステップS124)。

【0042】

視覚刺激制御部123は、記憶部122から出力情報 $Sig-1, \dots, Sig-N$ を読み込んで視覚刺激発生装置13-1, $\dots$, 13-Nにそれぞれ送る。各視覚刺激発生装置13-n($n=1, \dots, N$)は、送られた出力情報 $Sig-n$ に基づく視覚刺激パターン $VS(Sig-n)$ を提示(出力)する(ステップS123)。

40

【0043】

音 $SO(Info-1), \dots, SO(Info-N)$ および視覚刺激パターン $VS(Sig-1), \dots, VS(Sig-N)$ が提示された利用者10は、何れかの音 $SO(Info-n)$ に対して聴覚的に注意を向ける。例えば、利用者10は、前述の課題(タスク)を実行することで何れかの音 $SO(Info-n)$ に対して聴覚的に注意を向ける。瞳孔径取得装置15は、利用者10の瞳孔径 Pub を測定して、瞳孔径 Pub の時系列データを特徴量抽出部125に送る。

【0044】

特徴量抽出部125は、さらに記憶部122から視覚刺激パターン $VS(Sig-1), \dots, VS(Sig-N)$ に対応する出力情報 $Sig-1, \dots, Sig-N$ を抽出する。特徴量抽出部125は、瞳孔径 P

50

ubの時系列データおよび出力情報Sig-1,..., Sig-Nを用い、視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)のそれぞれと利用者10の瞳孔径変化量VPubとの相関の強さに基づく特徴量fを得て出力する。特徴量fの構成は、学習用音が音に置換され、学習用音源が音源に置換され、学習用視覚刺激パターンTVS(Sig-1), ..., TVS(Sig-N)が視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)に置換され、学習用利用者が利用者10に置換され、学習用利用者の学習用瞳孔径変化量TVPubが利用者10の瞳孔径変化量VPubに置換される以外、上述した学習用特徴量Tf_jの構成と同一である。例えば、特徴量抽出部125は、以下のように特徴量fを得る。

【0045】

(4.1) 特徴量抽出部125は、瞳孔径Pubの時系列データに対する前処理を行い、前処理後の瞳孔径Pub'の時系列データを得る。前処理としては、例えば、線形補間や2次スプライン補間などを用い、瞳孔径Pubの時系列データから利用者10の瞬き等によって欠損した部分を補間する処理を例示できる。さらに、視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)の明滅周波数に応じたローパスフィルタ（例えば、すべての明滅周波数を含む帯域を通すローパスフィルタ）を補完後の瞳孔径Pubの時系列データに適用し、ノイズの軽減を行ってもよい。

【0046】

(4.2) 特徴量抽出部125は、利用者10が何れかの学習用音TSO_nについて聴覚的に注意を向けているときの瞳孔径Pub'の時系列データから、聴覚的に注意を向ける前の瞳孔径Pub'の平均値を減算し、z値による標準化を行って瞳孔径変化量VPubの時系列データを得る。

【0047】

(4.3) 特徴量抽出部125は、瞳孔径変化量VPubの時系列データとN個の視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)とに基づいて特徴量fを得る。特徴量fは、N個の視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)のそれぞれと瞳孔径変化量VPubとの相関の強さに基づくものである。以下に特徴量fを例示する。

【0048】

(4.3.1) 複数の周波数領域視覚刺激信号FVS(Sig-1), ..., FVS(Sig-N)それぞれのピーク周波数および/またはピーク周波数の近傍における周波数領域瞳孔径変化量信号FVPubの大きさ（例えば、FVPubの大きさのピーク値）を表す特徴量fが用いられてもよい。ここで、複数の周波数領域視覚刺激信号FVS(Sig-1), ..., FVS(Sig-N)のそれぞれは、複数の視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)のそれぞれを表す時系列信号を周波数領域に変換して得られる信号である。また、周波数領域瞳孔径変化量信号FVPubは、瞳孔径変化量VPubを表す時系列信号を周波数領域に変換して得られる信号である。

【0049】

(4.3.2) 上述したピーク周波数および/またはピーク周波数の近傍における周波数領域瞳孔径変化量信号FVPubの大きさに加え、または、これに代えて、その他の情報が特徴量fに含まれてもよい。例えば、(1)複数の周波数領域視覚刺激信号FVS(Sig-1), ..., FVS(Sig-N)それぞれのピーク周波数の倍周波数またはそれらの近傍における周波数領域瞳孔径変化量信号FVPubの大きさ（例えば、大きさのピーク値）、(2)N個の視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)のそれぞれの位相変化と瞳孔径変化量VPubの位相変化の同期度、(3)N個の視覚刺激パターンVS(Sig-1), ..., VS(Sig-N)のそれぞれを表す時系列信号に対応する系列SS(Sig-1), ..., SS(Sig-N)のそれぞれと瞳孔径変化量VPubを表す時系列信号に対応する系列PSとに対する相互相関関数の各最大値CCFmax(SS(Sig-1), PS), ..., CCFmax(SS(Sig-N), PS)などが特徴量fに含まれてもよい。

【0050】

ここで、ピーク周波数の倍周波数および/またはその近傍における周波数領域瞳孔径変化量信号FVPubの大きさ（例えば、FVPubの大きさのピーク値）が大きいほど、当該ピーク周波数に対応する視覚刺激パターンVS(Sig-n)と瞳孔径変化量VPubとの相関が強い。また、瞳孔径変化量VPubとの同期度が高い位相変化を持つ視覚刺激パターンVS(Sig-n)ほど

10

20

30

40

50

、瞳孔径変化量 V_{Pub} との相関が強い。また、相互相関関数の最大値 $CCF_{max}(SS(Sig-n), PS)$ が大きい視覚刺激パターン $VS(Sig-n)$ ほど、瞳孔径変化量 V_{Pub} との相関が強い。特徴量 f は推定部 1 2 6 に送られる（ステップ S 1 2 5）。

【 0 0 5 1 】

推定部 1 2 6 は、記憶部 1 2 2 からモデルパラメータ θ を読み込む。推定部 1 2 6 は、特徴量 f を用い、モデルパラメータ θ によって特定される推定モデル推定モデル $M(\theta)$ に基づき、音について利用者 1 0 が聴覚的に注意を向けた先の推定結果 $E=M(\theta;f)$ を得て出力する。すなわち、推定部 1 2 6 は、モデルパラメータ θ によって特定される推定モデル推定モデル $M(\theta)$ に特徴量 f を適用し、特徴量 f に対応する推定結果 E を得て出力する。前述のように、推定結果 E は、音源（例えば、音源装置 1 4 - n ）を表す情報（例えば、インデックス）であってもよいし、音 SO_n に対応する視覚刺激パターン $VS(Sig-n)$ を表す情報（例えば、明滅周波数）であってもよいし、音 SO_n を表す情報であってもよいし、これらの方向を表す情報であってもよい。また、推定結果 E は、1 個の「利用者が聴覚的に注意を向けた先」を表していてもよいし、複数の「利用者が聴覚的に注意を向けた先」を表していてもよいし、「利用者が聴覚的に注意を向けた先」の確率を表していてもよい（ステップ S 1 2 6）。

10

【 0 0 5 2 】

[第 2 実施形態]

第 1 実施形態の視覚刺激パターン $VS(Sig-n)$ は、周期的に時間変化する視覚刺激のパターンであった。しかし、視覚刺激パターン $VS(Sig-n)$ が、非周期的に時間変化する刺激パターンであってもよい。以下では、第 1 実施形態との相違点を中心に説明し、第 1 実施形態と共通する事項については同じ参照番号を引用して説明を省略・簡略化する。

20

【 0 0 5 3 】

< 全体構成 >

図 1 に例示するように、本実施形態の聴覚注意状態推定システム 2 は、学習装置 2 1、聴覚注意状態推定装置 2 2、複数の視覚刺激発生装置 1 3 - 1, ..., 1 3 - N 、複数の音源装置 1 4 - 1, ..., 1 4 - N 、および瞳孔径取得装置 1 5 を有し、利用者 1 0 の瞳孔径変化量に基づいて、当該利用者 1 0 が聴覚的に注意を向けた先を推定する。

【 0 0 5 4 】

< 学習装置 2 1 >

図 2 に例示するように、本実施形態の学習装置 2 1 は、入力部 1 1 1、記憶部 1 1 2、学習部 2 1 3、出力部 1 1 4、および制御部 1 1 7 を有する。

30

【 0 0 5 5 】

< 聴覚注意状態推定装置 2 2 >

図 3 に例示するように、本実施形態の聴覚注意状態推定装置 2 2 は、入力部 1 2 1、記憶部 1 2 2、視覚刺激制御部 2 2 3、聴覚情報制御部 1 2 4、特徴量抽出部 2 2 5、推定部 2 2 6、および制御部 1 2 7 を有する。

【 0 0 5 6 】

< 学習装置 2 1 の処理 >

学習装置 2 1（図 2）の入力部 1 1 1 には、学習データ $T=\{(Tf_1, Ta_1), \dots, (Tf_j, Ta_j)\}$ が入力され、学習データ T は記憶部 1 1 2 に格納される。第 1 実施形態との相違点は、本実施形態の学習用視覚刺激パターン $TVS(Sig-1), \dots, TVS(Sig-N)$ が非周期的に時間変化する互いに異なる刺激パターンであり、本実施形態の学習用特徴量 Tf_j がこのような学習用視覚刺激パターン $TVS(Sig-1), \dots, TVS(Sig-N)$ のそれぞれと学習用瞳孔径変化量 TV_{Pub} との相関の強さに基づく点である。

40

【 0 0 5 7 】

学習用瞳孔径変化量 TV_{Pub} に限定はないが、例えば、 N 個の学習用視覚刺激パターン $TVS(Sig-1), \dots, TVS(Sig-N)$ それぞれを表す時系列信号に対応する系列 $TSS(Sig-1), \dots, TSS(Sig-N)$ それぞれと学習用瞳孔径変化量 TV_{Pub} を表す時系列信号に対応する系列 TPS とに対する相互相関関数の各最大値 $CCF_{max}(TSS(Sig-1), TPS), \dots, CCF_{max}(TSS(Sig-N), TPS)$ 、

50

、TPS)などが学習用特徴量 Tf_j に含まれる。その他、これに加え、または、これに代えて、 N 個の学習用視覚刺激パターン $TVS(\text{Sig-1}), \dots, TVS(\text{Sig-N})$ のそれぞれの位相変化と学習用瞳孔径変化量 $TVPub$ の位相変化の同期度などが学習用特徴量 Tf_j に含まれてもよい(ステップS 2 1 1)。

【0058】

学習部2 1 3は、記憶部1 1 2から読み込んだ学習データ T を用いた学習処理(機械学習)によって推定モデル $M(\theta)$ を得、当該推定モデル $M(\theta)$ を特定するモデルパラメータ θ を出力する。学習部2 1 3の処理の第1実施形態との相違点は学習データ T のみである。出力部1 1 4はモデルパラメータ θ を聴覚注意状態推定装置2 2に送る(ステップS 2 1 3)。

10

【0059】

<聴覚注意状態推定装置2 2の処理>

学習装置1 1から送られたモデルパラメータ θ は聴覚注意状態推定装置2 2(図3)の入力部1 2 1に入力され、記憶部1 2 2に格納される。また、記憶部1 2 2には、視覚刺激パターン $VS(\text{Sig-1}), \dots, VS(\text{Sig-N})$ を表す出力情報 $\text{Sig-1}, \dots, \text{Sig-N}$ が格納され、音 $SO(\text{Info-1}), \dots, SO(\text{Info-N})$ を表す出力情報 $\text{Info-1}, \dots, \text{Info-N}$ が格納される。視覚刺激パターン $VS(\text{Sig-1}), \dots, VS(\text{Sig-N})$ のそれぞれは、非周期的に時間変化する刺激パターンである(ステップS 2 2 2)。

【0060】

聴覚情報制御部1 2 4の処理は、第1実施形態と同じである(ステップS 1 2 4)。

20

【0061】

視覚刺激制御部2 2 3は、記憶部1 2 2から出力情報 $\text{Sig-1}, \dots, \text{Sig-N}$ を読み込んで視覚刺激発生装置1 3 - 1, $\dots$, 1 3 - Nにそれぞれ送る。各視覚刺激発生装置1 3 - n ($n = 1, \dots, N$)は、送られた出力情報 Sig-n に基づく視覚刺激パターン $VS(\text{Sig-n})$ を提示する。ただし、視覚刺激パターン $VS(\text{Sig-1}), \dots, VS(\text{Sig-N})$ のそれぞれは、非周期的に時間変化する刺激パターンである(ステップS 2 2 3)。

【0062】

音 $SO(\text{Info-1}), \dots, SO(\text{Info-N})$ および視覚刺激パターン $VS(\text{Sig-1}), \dots, VS(\text{Sig-N})$ が提示された利用者1 0は、何れかの音 $SO(\text{Info-n})$ に対して聴覚的に注意を向ける。瞳孔径取得装置1 5は、利用者1 0の瞳孔径 Pub を測定して、瞳孔径 Pub の時系列データを特徴量抽出部2 2 5に送る。

30

【0063】

特徴量抽出部2 2 5は、瞳孔径 Pub の時系列データおよび記憶部1 2 2から抽出した出力情報 $\text{Sig-1}, \dots, \text{Sig-N}$ を用い、覚刺激パターン $VS(\text{Sig-1}), \dots, VS(\text{Sig-N})$ のそれぞれと利用者1 0の瞳孔径変化量 V_{Pub} との相関の強さに基づく特徴量 f を得て出力する。特徴量 f の構成は、学習用音が音に置換され、学習用音源が音源に置換され、学習用視覚刺激パターン $TVS(\text{Sig-1}), \dots, TVS(\text{Sig-N})$ が視覚刺激パターン $VS(\text{Sig-1}), \dots, VS(\text{Sig-N})$ に置換され、学習用利用者が利用者1 0に置換され、学習用利用者の学習用瞳孔径変化量 $TVPub$ が利用者1 0の瞳孔径変化量 V_{Pub} に置換される以外、上述した学習用特徴量 Tf_j の構成と同一である。第1実施形態との相違点は、本実施形態の視覚刺激パターン $VS(\text{Sig-1}), \dots, VS(\text{Sig-N})$ が非周期的に時間変化する互いに異なる刺激パターンであり、本実施形態の特徴量 f がこのような視覚刺激パターン $VS(\text{Sig-1}), \dots, VS(\text{Sig-N})$ のそれぞれと瞳孔径変化量 V_{Pub} との相関の強さに基づく点である。例えば、特徴量抽出部2 2 5は、第1実施形態で説明した(4.1)および(4.2)の処理を実行し、(4.3)で N 個の視覚刺激パターン $VS(\text{Sig-1}), \dots, VS(\text{Sig-N})$ のそれぞれの位相変化と瞳孔径変化量 V_{Pub} の位相変化の同期度、 N 個の視覚刺激パターン $VS(\text{Sig-1}), \dots, VS(\text{Sig-N})$ のそれぞれを表す時系列信号に対応する系列 $SS(\text{Sig-1}), \dots, SS(\text{Sig-N})$ と瞳孔径変化量 V_{Pub} を表す時系列信号に対応する系列 PS とに対する相互相関関数の各最大値 $CCF_{\max}(SS(\text{Sig-1}), PS), \dots, CCF_{\max}(SS(\text{Sig-N}), PS)$ などを含む特徴量 f を得る。その他、これに加え、または、これに代えて、 N 個の視覚刺激パターン $VS(\text{Sig-1}), \dots, VS(\text{Sig-N})$ のそれぞれの位相変化と瞳孔径変化量 V_{Pub} の

40

50

位相変化の同期度などが特徴量 f に含まれてもよい。特徴量 f は推定部226に送られる（ステップS225）。

【0064】

推定部226は、記憶部122からモデルパラメータ θ を読み込む。推定部226は、特徴量 f を用い、モデルパラメータ θ によって特定される推定モデル推定モデル $M(\theta)$ に基づき、音について利用者10が聴覚的に注意を向けた先の推定結果 $E=M(\theta;f)$ を得て出力する（ステップS226）。

【0065】

[第1実施形態および第2実施形態の変形例1]

第1および第2実施形態では、音源装置14-nが n 番目の音 $SO(Info-n)$ を発する音源であった。しかしながら、複数の音源装置14-1, ..., 14-Nによって、各 n 番目（ $n=1, \dots, N$ ）の音 $SO(Info-n)$ が空間中の互いに異なる位置 y_n にそれぞれ定位されてもよい。この場合、空間中の各位置 y_n が n 番目の音 $SO(Info-n)$ を発する音源となり、視覚刺激発生装置13-nは音源である位置 y_n に対応する位置に配置される。例えば、視覚刺激発生装置13-nは位置 y_n または位置 y_n の近傍に配置される。

【0066】

[第3実施形態]

第1実施形態および第2実施形態ならびにそれらの変形例では、学習用視覚刺激パターンおよび視覚刺激パターンが、視覚刺激発生装置などの視覚刺激パターンを提示するための専用装置（以下、「視覚刺激専用装置」という）から提示された。しかし、視覚刺激専用装置以外の装置、風景、機械、植物、動物等の像の時間変化を学習用視覚刺激パターンおよび視覚刺激パターンとして利用してもよい。この場合、学習用視覚刺激パターンは、このような像をカメラ等で撮影することで得られた映像から生成される。また、聴覚注意状態の推定時における視覚刺激パターンは、利用者10が視覚刺激専用装置以外の装置、風景、機械、植物、動物等から視覚的に直接知覚するパターンとなる。この例の場合、視覚刺激制御部123, 223、および、視覚刺激発生装置13-1, ..., 13-Nは省略可能である。

【0067】

また、第1実施形態および第2実施形態ならびにそれらの変形例では、学習用音および音が、音源装置などの専用装置（以下、「音提示専用装置」という）から提示された。しかし、音提示専用装置以外の装置、風景、機械、植物、動物等から発せられる音を利用してもよい。学習用音は、このような音をマイクロホン等で録音して得られたオーディオ信号から生成できる。また、聴覚注意状態の推定時に利用者10に提示される音は、利用者10が音提示専用装置以外の装置、風景、機械、植物、動物等から聴覚的に直接知覚する音となる。この例の場合、聴覚情報制御部124、および、音源装置14-1, ..., 14-Nは省略可能である。

【0068】

その他は、第1実施形態および第2実施形態で説明した通りである。

【0069】

[ハードウェア構成]

各実施形態における学習装置11, 21および聴覚注意状態推定装置12, 22は、例えば、CPU (central processing unit) 等のプロセッサ (ハードウェア・プロセッサ) やRAM (random-access memory) ・ROM (read-only memory) 等のメモリ等を備える汎用または専用のコンピュータが所定のプログラムを実行することで構成される装置である。すなわち、各実施形態における学習装置11, 21および聴覚注意状態推定装置12, 22は、例えば、それぞれが有する各部を実装するように構成された処理回路 (processing circuitry) を有する。このコンピュータは1個のプロセッサやメモリを備えていてもよいし、複数個のプロセッサやメモリを備えていてもよい。このプログラムはコンピュータにインストールされてもよいし、予めROM等に記録されていてもよい。また、CPUのようにプログラムが読み込まれることで機能構成を実現する電子回路 (circu

10

20

30

40

50

itry)ではなく、単独で処理機能を実現する電子回路を用いて一部またはすべての処理部が構成されてもよい。また、1個の装置を構成する電子回路が複数のCPUを含んでいてもよい。

【0070】

図6は、各実施形態における学習装置11, 21および聴覚注意状態推定装置12, 22のハードウェア構成を例示したブロック図である。図6に例示するように、この例の学習装置11, 21および聴覚注意状態推定装置12, 22は、CPU (Central Processing Unit) 10a、入力部10b、出力部10c、RAM (Random Access Memory) 10d、ROM (Read Only Memory) 10e、補助記憶装置10f及びバス10gを有している。この例のCPU 10aは、制御部10aa、演算部10ab及びレジスタ10acを有し、レジスタ10acに読み込まれた各種プログラムに従って様々な演算処理を実行する。また、入力部10bは、データが入力される入力端子、キーボード、マウス、タッチパネル等である。また、出力部10cは、データが出力される出力端子、ディスプレイ、所定のプログラムを読み込んだCPU 10aによって制御されるLANカード等である。また、RAM 10dは、SRAM (Static Random Access Memory)、DRAM (Dynamic Random Access Memory)等であり、所定のプログラムが格納されるプログラム領域10da及び各種データが格納されるデータ領域10dbを有している。また、補助記憶装置10fは、例えば、ハードディスク、MO (Magneto-Optical disc)、半導体メモリ等であり、所定のプログラムが格納されるプログラム領域10fa及び各種データが格納されるデータ領域10fbを有している。また、バス10gは、CPU 10a、入力部10b、出力部10c、RAM 10d、ROM 10e及び補助記憶装置10fを、情報のやり取りが可能なように接続する。CPU 10aは、読み込まれたOS (Operating System) プログラムに従い、補助記憶装置10fのプログラム領域10faに格納されているプログラムをRAM 10dのプログラム領域10daに書き込む。同様にCPU 10aは、補助記憶装置10fのデータ領域10fbに格納されている各種データを、RAM 10dのデータ領域10dbに書き込む。そして、このプログラムやデータが書き込まれたRAM 10d上のアドレスがCPU 10aのレジスタ10acに格納される。CPU 10aの制御部10aaは、レジスタ10acに格納されたこれらのアドレスを順次読み出し、読み出したアドレスが示すRAM 10d上の領域からプログラムやデータを読み出し、そのプログラムが示す演算を演算部10abに順次実行させ、その演算結果をレジスタ10acに格納していく。このような構成により、学習装置11, 21および聴覚注意状態推定装置12, 22の機能構成が実現される。

【0071】

上述のプログラムは、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。

【0072】

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。上述のように、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。そして、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。また、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。また、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ASP（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。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの（コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等）を含むものとする。

【0073】

各実施形態では、コンピュータ上で所定のプログラムを実行させることにより、本装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

10

【0074】

なお、本発明は上述の実施形態に限定されるものではない。例えば、上述の各実施形態では、学習によって推定モデルを得、当該推定モデルを用い、利用者が聴覚的に注意を向けた先を推定した。しかしながら、互いに異なる複数の音源に対応する互いに異なる複数の視覚刺激パターンのそれぞれと利用者の瞳孔径変化量との相関の強さに基づく特徴量を用いるのであれば、どのような方法で利用者が聴覚的に注意を向けた先を推定してもよい。例えば、過去に得られた特徴量のサンプリングによって閾値を定め、当該閾値と新たに得た特徴量との比較によって、利用者が聴覚的に注意を向けた先を推定してもよい。

【0075】

また、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に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。

20

【符号の説明】

【0076】

- 1, 2 聴覚注意状態推定システム
- 11, 21 学習装置
- 113, 213 学習部
- 12, 22 聴覚注意状態推定装置
- 126, 226 推定部

30

40

50

【図面】

【図 1】

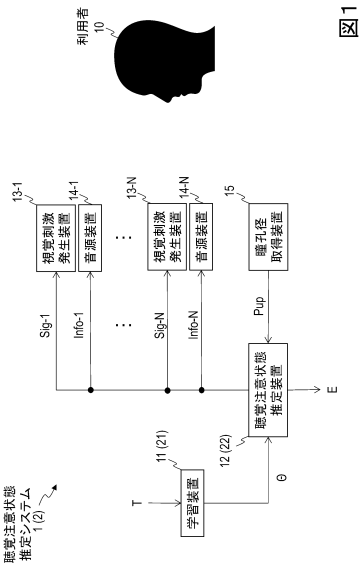


図 1

【図 2】

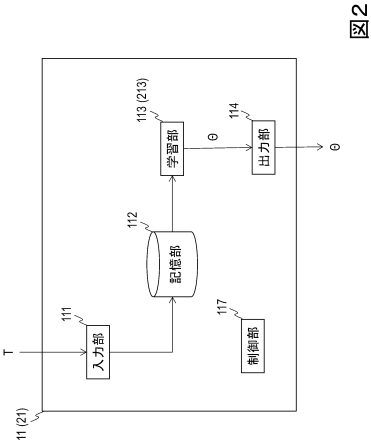


図 2

【図 3】

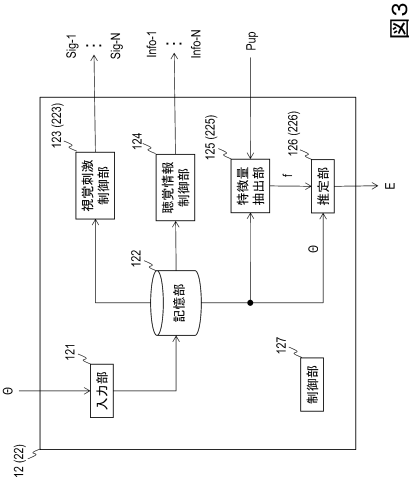


図 3

【図 4】

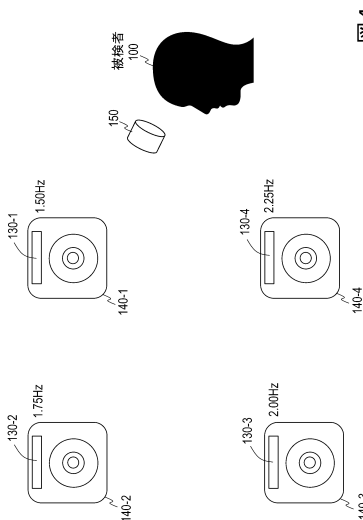


図 4

【図 5】

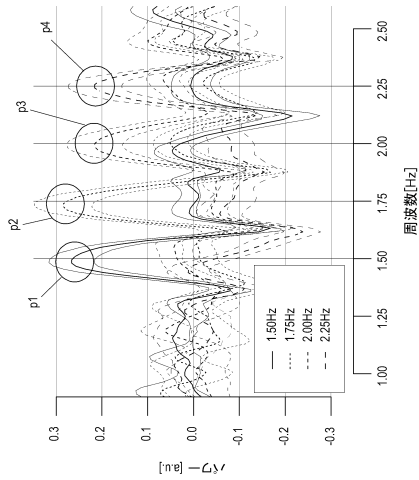


図5

【図 6】

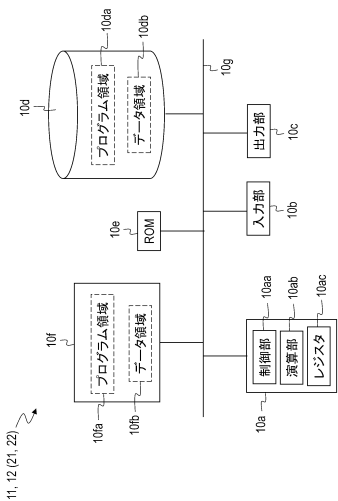


図6

10

20

30

40

50

フロントページの続き

- (72)発明者 古川 茂人
東京都千代田区大手町一丁目 5 番 1 号 日本電信電話株式会社内
- (72)発明者 楊 詠皓
東京都千代田区大手町一丁目 5 番 1 号 日本電信電話株式会社内
- 審査官 後藤 昌夫
- (56)参考文献 特開 2 0 1 9 - 1 2 6 4 2 3 (J P , A)
米国特許出願公開第 2 0 2 0 / 0 2 5 3 5 2 6 (U S , A 1)
特開 2 0 2 1 - 0 1 9 9 6 3 (J P , A)
特開 2 0 1 5 - 1 3 2 7 8 3 (J P , A)
特開 2 0 1 0 - 1 0 4 7 5 4 (J P , A)
国際公開第 2 0 1 3 / 0 2 3 0 5 6 (W O , A 1)
- (58)調査した分野 (Int.Cl. , D B 名)
A 6 1 B 3 / 0 0 - 3 / 1 8