(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2020年8月27日(27.08.2020)

WIPO | PCT

(10) 国際公開番号
WO 2020/171049 A1

(51) 国際特許分類:
G10L 19/002 (2013.01) *G10L 19/035* (2013.01)
G10L 19/008 (2013.01)

(21) 国際出願番号： PCT/JP2020/006211

(22) 国際出願日： 2020年2月18日(18.02.2020)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

(30) 優先権データ：
特願 2019-027035 2019年2月19日(19.02.2019) JP

(71) 出願人: 公立大学法人秋田県立大学(**AKITA PREFECTURAL UNIVERSITY**) [JP/JP]; 〒0100195 秋田県秋田市下新城中野字街道端西２４１－４３８ Akita (JP).

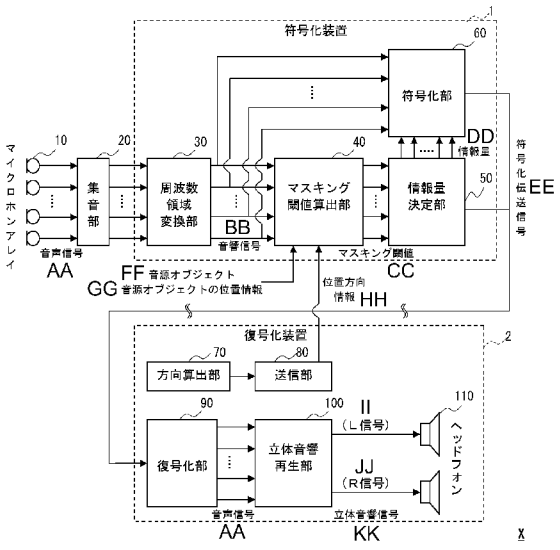
(72) 発明者: 西口 正之 (**NISHIGUCHI, Masayuki**); 〒0150055 秋田県由利本荘市土谷字海老ノ口８４－４ 公立大学法人秋田県立大学 システム科学技術学部内 Akita (JP). 加藤 巧大(**KATO, Kodai**); 〒3114152 茨城県水戸市河和田１丁目１９１１－１９ クレア２０２号 Ibaraki (JP).

(74) 代理人: 堀 城之, 外 (**HORI, Shiroyuki** et al.); 〒1000013 東京都千代田区霞が関３－３－１ 尚友会館１階 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH,

(54) **Title:** ACOUSTIC SIGNAL ENCODING METHOD, ACOUSTIC SIGNAL DECODING METHOD, PROGRAM, ENCODING DEVICE, ACOUSTIC SYSTEM AND COMPLEXING DEVICE

(54) 発明の名称： 音響信号符号化方法、音響信号復号化方法、プログラム、符号化装置、音響システム、及び復号化装置



1...Encoding device
2...Decoding device
10...Microphone array
20...Sound collection unit
30...Frequency domain conversion unit
40...Masking threshold calculation unit
50...Information amount determination unit
60...Encoding unit
70...Direction calculation unit
80...Transmission unit
90...Decoding unit
100...Stereophonic reproduction unit
110...Headphones
AA...Sound signal
BB...Acoustic signal
CC...Masking threshold
DD...Amount of information
EE...Encoded transmission signal
FF...Sound source object
GG...Sound source object position information
HH...Position/direction information
II...L signal
JJ...R signal
KK...Stereophonic signal

(57) **Abstract:** Provided is an acoustic signal encoding method capable of encoding even acoustic signals that have many channels at a sufficient bitrate. The acoustic signal encoding method is implemented by an encoding device 1, and encodes a multichannel acoustic signal. First, a masking threshold corresponding to the spatial auditory masking effect is calculated. Then, the amount of information to be allocated for each channel of the multichannel acoustic signal is determined using the calculated masking threshold. Thereafter, the multichannel acoustic signal is encoded using the respectively assigned amounts of information. Thus, it is possible to encode even multichannel acoustic signals at a sufficient bitrate.

WO 2020/171049 A1

［続葉有］

KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告(条約第21条(3))

— 請求の範囲の補正の期限前の公開であり、補正を受理した際には再公開される。(規則48.2(h))

(57) 要約:チャンネル数が多い音響信号でも十分なビットレートで符号化が可能な音響信号符号化方法を提供する。この音響信号符号化方法では、符号化装置1により実行される、複数のチャンネルの音響信号を符号化する。まず、聴覚の空間的マスキング効果に対応したマスキング閾値を算出する。そして、算出されたマスキング閾値により、複数チャンネルの音響信号を各チャンネルに割り振る情報量を決定する。この上で、複数のチャンネルの音響信号を、それぞれ割り振られた情報量で符号化する。これにより、複数のチャンネルの音響信号でも十分なビットレートでの符号化も可能となる。

明　細　書

発明の名称 ： 音響信号符号化方法、音響信号復号化方法、プログラム、符号化装置、音響システム、及び復号化装置

技術分野

[0001] 本発明は、特に音響信号符号化方法、音響信号復号化方法、プログラム、符号化装置、音響システム、及び複合化装置に関する。

背景技術

[0002] 従来、音響信号（オーディオ信号）の符号化においては、複数のチャンネルに入力した音響信号のチャンネル毎の量子化におけるビット数を時間軸又は周波数軸で適応的に割り当てるビットアロケーション（ビット割り当て）による音響符号化技術がある。

近年、標準的に使用されているＭＰＥＧ－２　ＡＡＣ、ＭＰＥＧ－４　ＡＡＣ、ＭＰ３等の音響信号の符号化においては、このビット割り当てにおいて、周波数軸における聴覚のマスキング効果が利用されている。

[0003] この聴覚におけるマスキング効果とは、ある音が他の音の存在によって聴こえにくくなる効果である。

特許文献１には、聴覚のマスキング効果が利用された音響信号符号化の技術の一例が記載されている。特許文献１の技術では、聴覚のマスキング効果を利用するために、マスキング効果のビット割り当ての閾値（以下、マスキング閾値という。）が計算されている。

先行技術文献

特許文献

[0004] 特許文献1：特開平５－２４８９７２号公報

非特許文献

[0005] 非特許文献1：Ａｎｄｒｅａｓ　Ｓｐａｎｉａｓ他著、「Ａｕｄｉｏ　Ｓｉｇａｌ　Ｐｒｏｃｅｓｓｉｎｇ　ａｎｄ　Ｃｏｄｉｎｇ」、米国、、Ｗｉｌｅ

ｙ－Ｉｎｔｅｒｓｃｉｅｎｃｅ，Ｊｏｈｎ Ｗｉｌｅｙ ＆ Ｓｏｎｓ，Ｉｎｃ、２００７年

発明の概要

発明が解決しようとする課題

[0006] しかし、従来のマスキング閾値の計算では、複数のチャンネル同士の空間的な関係は考慮されていなかった為、チャンネル数が多い音響信号ではビットレート（帯域）が不足するおそれが生じるという問題があった。

[0007] 本発明は、このような状況に鑑みてなされたものであり、上述の問題を解消することを目的とする。

課題を解決するための手段

[0008] 本発明の音響信号符号化方法は、符号化装置により実行される、複数のチャンネルの音響信号を符号化する音響信号符号化方法であって、聴覚の空間的マスキング効果に対応したマスキング閾値を算出し、算出された前記マスキング閾値により、各前記チャンネルに割り振る情報量を決定し、複数の前記チャンネルの音響信号を、それぞれ割り振られた前記情報量で符号化することを特徴とする。

本発明のプログラムは、符号化装置により実行される、複数のチャンネルの音響信号を符号化するプログラムであって、前記符号化装置に、聴覚の空間的マスキング効果に対応したマスキング閾値を算出させ、算出された前記マスキング閾値により、各前記チャンネルに割り振る情報量を決定させ、複数の前記チャンネルの音響信号を、それぞれ割り振られた前記情報量で符号化させることを特徴とする。

本発明の符号化装置は、複数のチャンネルの音響信号、及び／又は、音源オブジェクト並びに該音源オブジェクトの位置情報を符号化する符号化装置であって、聴覚の空間的マスキング効果に対応したマスキング閾値を算出するマスキング閾値算出部と、前記マスキング閾値算出部により算出された前記マスキング閾値により、各前記チャンネル及び／又は前記音源オブジェクトに割り振る情報量を決定する情報量決定部と、複数の前記チャンネルの音

響信号、及び／又は、前記音源オブジェクト並びに前記音源オブジェクトの位置情報を、それぞれ割り振られた前記情報量で符号化する符号化部とを備えることを特徴とする。

本発明の音響システムは、前記符号化装置と、復号化装置とを備えた音響システムであって、前記復号化装置は、受聴者の向いている方向を算出する方向算出部と、前記方向算出部により算出された前記方向を前記符号化装置に送信する送信部と、前記符号化装置で符号化された複数の前記チャンネルの音響信号、及び／又は前記音源オブジェクトを音声信号に復号化する復号化部を備え、前記符号化装置の前記マスキング閾値算出部は、前記マスキング閾値を、前記受聴者の位置と前記方向に対する、各前記チャンネル間及び／又は各前記音源オブジェクト間の空間的距離及び／又は方向に基づいた前記空間的マスキング効果に対応して算出することを特徴とする。

本発明の復号化装置は、聴覚の空間的マスキング効果に対応したマスキング閾値により、各チャンネル及び／又は音源オブジェクトに割り振る情報量が決定され、複数の前記チャンネルの音響信号、及び／又は、前記音源オブジェクト並びに前記音源オブジェクトの位置情報を、それぞれ割り振られた前記情報量で符号化された信号を取得する信号取得部と、前記信号取得部により取得された信号から、符号化された複数の前記チャンネルの音響信号、及び／又は前記音源オブジェクトを音声信号に復号化する復号化部とを備えることを特徴とする。

発明の効果

[0009] 本発明によれば、聴覚の空間的マスキング効果に対応したマスキング閾値を算出し、算出されたマスキング閾値により、複数チャンネルの音響信号を各前記チャンネルに割り振る情報量を決定し、割り振られた情報量で符号化することで、チャンネル数が多い音響信号でも十分なビットレートでの符号化が可能な音響信号符号化方法を提供することができる。

図面の簡単な説明

[0010] [図1]本発明の実施の形態に係る音響システムのシステム構成図である。

[図2]本発明の実施の形態に係る音響符号化復号化処理のフローチャートである。

[図3]図２に示す音響符号化復号化処理の概念図である。

[図4]図２に示す音響符号化復号化処理の概念図である。

[図5]本発明の実施例に係る聴取実験の測定システムを示す概念図である。

[図6]本発明の実施例に係る聴取実験における閾値探索を示す概念図である。

[図7]本発明の実施例に係る聴取実験における回答画面の画面例である。

[図8]本発明の実施例に係るマスカーの方位が０°の際のマスキング閾値のピーク値を、横軸をマスキーの方位としてプロットしたグラフである。

[図9]本発明の実施例に係るマスカーの方位が４５°の際のマスキング閾値のピーク値を、横軸をマスキーの方位としてプロットしたグラフである。

[図10]本発明の実施例に係るマスカーの方位が９０°の際のマスキング閾値のピーク値を、横軸をマスキーの方位としてプロットしたグラフである。

[図11]本発明の実施例に係るマスカーの方位が１３５°の際のマスキング閾値のピーク値を、横軸をマスキーの方位としてプロットしたグラフである。

発明を実施するための形態

[0011] ＜実施の形態＞

〔音響システムＸの制御構成〕

まず、図１を参照して、本発明の実施の形態に係る音響システムＸの制御構成について説明する。

音響システムＸは、複数のチャンネルの音響信号を取得し、符号化装置１により符号化し、伝送し、復号化装置２により復号化し、再生することが可能なシステムである。

[0012] 符号化装置１は、音響信号を符号化する装置である。本実施形態において、符号化装置１は、例えば、ＰＣ（Personal Computer）、サーバー、これらに装着するエンコーダーボード、専用のエンコーダー等である。本実施形態の符号化装置１は、複数のチャンネルの音響信号、及び／又は、音源オブジェクト並びに該音源オブジェクトの位置情報を符号化する。たとえば、符号

化装置１は、ＭＰＥＧ－２　ＡＡＣ、ＭＰＥＧ－４　ＡＡＣ、ＭＰ３、Ｄｏｌｂｙ（登録商標）Ｄｉｇｉｔａｌ、ＤＴＳ（登録商標）等の音響符号化の方式に対応して、２チャンネル、５．１チャンネル、７．１チャンネル、２２．２チャンネル等の複数チャンネルの音響信号についての符号化を行う。

[0013] 復号化装置２は、復号化装置２により符号化された音響信号を復号化する装置である。本実施形態において、復号化装置２は、例えば、ＶＲ（Virtual Reality）やＡＲ（Augmented Reality）用のＨＭＤ（Head-Mounted Display）、スマートフォン（Smart Phone）、ゲーム専用機、家庭用テレビ、無線接続ヘッドフォン、仮想多チャンネルヘッドフォン、映画館やパブリックビューイング会場の機器、専用のデコーダー及びヘッドトラッキングセンサー等である。復号化装置２は、符号化装置１で符号化され、有線や無線で伝送された音響信号を復号化して、再生する。

[0014] 音響システムＸは、主に、マイクロホンアレイ１０、集音部２０、周波数領域変換部３０、マスキング閾値算出部４０、情報量決定部５０、符号化部６０、方向算出部７０、送信部８０、復号化部９０、立体音響再生部１００、及びヘッドフォン１１０を含んで構成される。

[0015] このうち、周波数領域変換部３０、マスキング閾値算出部４０、情報量決定部５０、及び符号化部６０は、本実施形態の符号化装置１（送信側）として機能する。

方向算出部７０、送信部８０、復号化部９０、立体音響再生部１００、及びヘッドフォン１１０は、本実施形態の復号化装置２（受信側）として機能する。

[0016] マイクロホンアレイ１０は、様々な音が様々な場所に存在するような空間である音空間の音声を収音する。具体的には、例えば、マイクロホンアレイ１０は、３６０°の複数方向の音波を取得する。この際、ビームフォーミング処理によって指向性を制御し、各方向にビームを向けることで、音空間の空間サンプリングを行い、多チャンネルの音声ビーム信号を取得することが可能である。具体的には、本実施形態のビームフォーミングでは、マイクロ

ホンアレイ１０の各マイクロホンに到来する音波の位相差をフィルターにより制御し、各マイクロホンに到来する方向の信号を強調する。この上で、空間サンプリングとして、音場を空間的に切り分けて、空間的情報を含めたまま、多チャンネルで集音する。

[0017] 集音部２０は、複数のチャンネルの音声をまとめて、音響信号として符号化装置１に送信するミキサー等のデバイスである。

[0018] 周波数領域変換部３０は、空間サンプリングすることで得られた方向別の音声ビーム信号を数マイクロ秒～数十ミリ秒程度のウィンドウ（フレーム）に切り出し、ＤＦＴ（discrete Fourier transformation、離散フーリエ変換）やＭＤＣＴ（Modified Discrete Cosine Transform、変形離散コサイン変換）等によって、時間領域から周波数領域へ変換する。このフレームは、例えば、サンプリング周波数４８ｋＨｚ、量子化ビット数１６ビットで、２０４８サンプル程度を用いることが好適である。周波数領域変換部３０は、このフレームを、各チャンネルの音響信号として出力する。すなわち、本実施形態の音響信号は、周波数領域の信号となる。

[0019] マスキング閾値算出部４０は、周波数領域変換部３０により変換された各チャンネルの音響信号から、聴覚の空間的マスキング効果に対応したマスキング閾値を算出する。この際、マスキング閾値算出部４０は、空間的マスキング効果を考慮したモデルを適用して、その上で、周波数領域でのマスキング閾値を計算する。この周波数領域でのマスキング閾値の計算自体は、例えば、非特許文献１に記載の方式で実現することが可能である。

[0020] または、マスキング閾値算出部４０は、音源オブジェクトを取得し、同様に、聴覚の空間的マスキング効果に対応したマスキング閾値を算出することも可能である。この音源オブジェクトは、空間的に異なる位置から発生された複数の音響信号のそれぞれを示す。この音源オブジェクトは、例えば、位置情報が付された音響信号である。これは、例えば、オーケストラの各楽器を収録するようなマイクの出力信号、ゲーム等で用いるサンプリングされた音声信号等が、周波数領域の音響信号に変換されたものでもよい。

さらに、マスキング閾値算出部４０は、一旦、集音され、フラッシュメモリー、ＨＤＤ、光学記録媒体等の記録媒体に格納された音響信号を取得したり、変換したりして、周波数マスキングを計算することも可能である。

[0021] 具体的には、上述の空間的マスキング効果のモデルとして、マスキング閾値算出部４０は、マスキング閾値を、受聴者の位置方向情報に対する、各チャンネル間及び／又は各音源オブジェクト間の空間的距離及び／又は方向に基づいた空間的マスキング効果に対応して算出することも可能である。

または、マスキング閾値算出部４０は、マスキング閾値を、各チャンネル間及び／又は各音源オブジェクト間の空間的距離及び／又は方向に基づいた空間的マスキング効果に対応して算出してもよい。

より具体的には、マスキング閾値算出部４０は、マスキング閾値を、チャンネル及び／又は音源オブジェクト間の空間的距離及び／又は方向が近づくほど相互に及ぼす影響が大きくなり、離れるほど相互に及ぼす影響が小さくなるような空間的マスキング効果に対応して算出してもよい。

加えて、マスキング閾値算出部４０は、マスキング閾値を、受聴者からみて前後対称の位置にあるチャンネル及び／又は音源オブジェクトについては、音源オブジェクト間の空間的距離及び／又は方向についての相互に及ぼす影響の度合いを変化させるような空間的マスキング効果に対応して算出してもよい。

さらに、マスキング閾値算出部４０は、マスキング閾値を、受聴者からみて後方の位置にあるチャンネル及び／又は音源オブジェクトについては、前後対称の位置に該当する前方に当該チャンネル及び／又は当該オブジェクトが存在するような空間的マスキング効果に対応して算出してもよい。

[0022] 具体的には、マスキング閾値算出部４０は、マスキング閾値を算出する際、

下記の式（１）で調整してもよい。

$$T = \beta \{ max（y1, \alpha y2）- 1 \}$$

$$y1 = f（x - \theta）$$

ｙ２＝ｆ（１８０－ｘ－θ）　　……　式（１）

ただし、Ｔは前記マスキング閾値を算出するために、各チャンネル信号の周波数領域におけるマスキング閾値に乗ずる重み、θはマスカーの方位、αはマスカーの周波数で制御される定数、βはマスカーの信号がトーン性の信号かノイズ性の信号かに対応して制御される定数、ｘは求める方向又はマスキーの方位を示す。

[0023] より具体的に説明すると、本実施形態において、聴こえを妨害する音を「マスカー」といい、聴こえが妨害される音を「マスキー」という。ｍａｘは、引数内の最大値を返す関数である。定数については、マスカーが４００Ｈｚの場合、α＝１、マスカーが１ｋＨｚの場合、α＝０．８のような値を用いることが可能である。マスカーがノイズ性の場合は、β＝１１～１４、純音（トーン性）の場合は３～５程度の値を用いることが可能である。すなわち、マスカーがトーン性の場合は、Ｔは、ｘの値にかかわらず、全てのθについてフラットとなる。

[0024] この式（１）のｆ（ｘ）は、例えば、下記の式（２）に示す三角波のようなリニアな関数を用いることが可能である。

[0025] [数1]

$$f(x)=\begin{cases}-\dfrac{1}{90}x+1 & (0^\circ \le x \le 180^\circ)\\ \dfrac{1}{90}x-2 & (180^\circ \le x \le 360^\circ)\end{cases}$$

……　式（２）

[0026] このうち、ｘは、求める方位、又は、マスキーの方位を用いることが可能である。この方位は、マイクロホンのビームフォーミングの方向、音源オブジェクトの方向等に対応する。

なお、ｆ（ｘ）として、ｆ（ｘ）＝ｃｏｓ（ｘ）のような式も、用いることが可能である。さらに、ｆ（ｘ）として、これ以外の、例えば、実際のマスカー、マスキーの実験結果から算出された関数等も用いることが可能である。

[0027] マスキング閾値算出部４０は、マスキング閾値を、各チャンネル及び／又は音源オブジェクトの信号が、トーン性の信号かノイズ性の信号かに対応して、各チャンネル及び／又は音源オブジェクトの信号の相互に及ぼす影響の度合いを変化させる空間的マスキング効果に対応して算出してもよい。

[0028] 情報量決定部５０は、マスキング閾値算出部４０により算出されたマスキング閾値により、音源オブジェクトに割り振る情報量を決定する。本実施形態では、この情報量として、マスキング閾値に基づいた各音響信号のビット割り当てが行われる。情報量決定部５０は、このビット割り当てとして、Ｐｅｒｃｅｐｔｕａｌ　Ｅｎｔｒｏｐｙ（以下、「ＰＥ」という。）により、一サンプル当たりの平均ビット数を、マスキング閾値算出部４０により算出されたマスキング閾値に対応して算出することが可能である。

[0029] 符号化部６０は、複数のチャンネルの音響信号、及び／又は、音源オブジェクト並びに音源オブジェクトの位置情報を、それぞれ割り振られた情報量で符号化する。本実施形態では、符号化部６０は、情報量決定部５０により割り当てられたビット数に基づいて各音響信号を量子化し、伝送路へ送信する。この伝送路は、例えば、Ｂｌｕｅｔｏｏｔｈ（登録商標）、ＨＤＭＩ（登録商標）、ＷｉＦｉ、ＵＳＢ（Universal Serial Bus）、その他の有線や無線の情報伝送手段を用いることが可能である。より具体的には、インターネットやＷｉＦｉ等のネットワークを介した、ピアツーピア（Peer to Peer）通信によって伝送可能である。

[0030] 方向算出部７０は、受聴者の向いている方向を算出する。方向算出部７０は、例えば、ヘッドトラッキングが可能な加速度センサー、ジャイロセンサー、地磁気センサー等と、これらの出力を方向情報に変換する回路とを含む。

この上で、方向算出部７０は、算出された方向情報に、受聴者に対する音源オブジェクトや複数チャンネルの音響信号についての位置の関係を考慮した位置情報を加えた位置方向情報を算出可能である。

[0031] 送信部８０は、方向算出部７０により算出された位置方向情報を符号化装

置１に送信する。送信部８０は、例えば、音響信号の伝送路と同様の有線や無線の伝送により、位置方向情報をマスキング閾値算出部４０で受信可能に送出することが可能である。

[0032] 復号化部９０は、符号化装置１で符号化された複数のチャンネルの音響信号、及び／又は音源オブジェクトを音声信号に復号化する。復号化部９０は、例えば、まず、伝送路から受信した信号を逆量子化する。次に、ＩＤＦＴ（Inverse Discrete Fourier Transform、逆離散フーリエ変換、離散フーリエ逆変換）、ＩＭＤＣＴ（Inverse Modified Discrete Cosine Transform、逆変形離散コサイン変換）等により、周波数領域の信号を時間領域に戻して、各チャンネルの音声信号に変換する。

[0033] 立体音響再生部１００は、復号化部９０により復号化された音声信号を、受聴者に対する立体音響を再生するような立体音響信号に変換する。具体的には、立体音響再生部１００は、時間領域に戻された方向別のビーム信号をその方向にある音源から発せられた信号とみなして、ビーム方向のＨＲＴＦ（Head-Related Transfer Function、頭部伝達関数）をそれぞれ畳み込む。ＨＲＴＦは、耳殻、人頭及び肩までふくめた周辺物によって生じる音の変化を伝達関数として表現したものである。

次に、ＨＲＴＦが畳み込まれた信号にビーム方向別の重み付けを行ってから加算することで、聴取者に提示する２チャンネルの両耳信号を生成する。このうち、ビーム方向別重み付けとは、Ｌ信号及びＲ信号である両耳信号が再現したい音空間における両耳信号により近づくような重み付けを行う処理である。具体的には、ある音空間に存在する各音源に音源方向のＨＲＴＦをそれぞれ畳み込んで加算することにより、両耳信号を生成する。その両耳信号を目標信号とし、出力として得られた両耳信号が目標信号と等しくなるように、出力信号に重みを付加する処理を行う。

立体音響再生部１００は、上述のマスキング閾値とは別に、方向算出部７０により算出された位置方向情報により、ＨＲＴＦをアップデートし、立体音響を再生することが可能である。

[0034] ヘッドフォン１１０は、復号化され、立体音響化された音響を受聴者が再生するデバイスである。ヘッドフォン１１０は、Ｄ／Ａコンバーター、アンプ（Amplifier）、電磁ドライバー、ユーザーの装着する耳当て等を備えている。

[0035] これに加え、符号化装置１及び復号化装置２は、例えば、各種回路として、ＡＳＩＣ（Application Specific Processor、特定用途向けプロセッサー）、ＤＳＰ（Digital Signal Processor）、ＣＰＵ（Central Processing Unit、中央処理装置）、ＭＰＵ（Micro Processing Unit）、ＧＰＵ（Graphics Processing Unit）等の制御演算手段である制御部を含んでいる。

加えて、符号化装置１及び復号化装置２は、記憶手段として、ＲＯＭ（Read Only Memory）、ＲＡＭ（Random Access Memory）等の半導体メモリー、ＨＤＤ（Hard Disk Drive）等の磁気記録媒体、光学記録媒体等である記憶部を含んでいる。この記憶部には、本発明の実施の形態に係る各方法を実現するための制御プログラムが格納されている。

さらに、符号化装置１及び復号化装置２は、液晶ディスプレイや有機ＥＬディスプレイ等の表示手段、キーボード、マウスやタッチパネル等のポインティングデバイス等の入力手段、ＬＡＮボード、無線ＬＡＮボード、シリアル、パラレル、ＵＳＢ（Universal Serial Bus）等のインターフェイスを含んでいてもよい。

[0036] また、符号化装置１及び復号化装置２は、主に記憶手段に格納された各種プログラムを用いて制御部が実行することで、本発明の実施の形態に係る各方法を、ハードウェア資源を用いて実現することができる。

なお、上述の構成の一部又は任意の組み合わせをＩＣやプログラマブルロジックやＦＰＧＡ（Field-Programmable Gate Array）等でハードウェア的、回路的に構成してもよい。

[0037]〔音響システムＸによる音響符号化復号化処理〕

次に、図２及び図３を参照して、本発明の実施の形態に係る音響システムＸによる音響信号符号化復号化処理の説明を行う。

本実施形態の音響信号符号化復号化処理は、主に符号化装置１及び復号化装置２において、それぞれ、制御部が記憶部に格納された制御プログラムを、各部と協働し、ハードウェア資源を用いて制御して実行し、又は、各回路で直接実行する。

以下で、図２のフローチャートを参照して、音響信号符号化復号化処理の詳細をステップ毎に説明する。

[0038]（ステップＳ１０１）

まず、符号化装置１の周波数領域変換部３０が、音声データ取得処理を行う。

ここでは、集音者がスタジアム等に赴き、マイクロホンアレイ１０を用いて収音を行う。これにより、マイクロホンアレイ１０を中心とした各方向（θ）の音声信号が取得される。この際に、収音側では、「空間サンプリング」の考え方に基づいて収音を行う。空間サンプリングは、音場を空間的に切り分けて多チャンネルで収音するものである。本実施形態では、例えば、左右０°～３６０°を区切った特定ステップの音声信号を、複数チャンネルに対応して収音する。なお、上下方向の０°～３６０°についても、特定ステップに区切って収音することが可能である。

周波数領域変換部３０は、これらの集音された音声データ等を切り出し、ＤＦＴ、ＭＤＣＴ等によって、時間領域から周波数領域の信号へ変換し、音響信号として記憶部に格納する。

[0039]（ステップＳ２０１）

ここで、復号化装置２の方向算出部７０が、方向算出処理を行う。

方向算出部７０は、受聴者の向いている方向情報と、音響データに対しての位置情報とを算出する。

[0040]（ステップＳ２０２）

次に、送信部８０が、方向送信処理を行う。

送信部８０は、方向算出部７０により算出された位置方向情報を、符号化装置１へ送信する。

[0041]（ステップＳ１０２）

ここで、符号化装置１のマスキング閾値算出部４０が、マスキング閾値算出処理を行う。本実施形態では、周波数領域でマスキング閾値Ｔを計算して、後述する空間的マスキングのマスキング閾値を更に算出し、ビット割り当てを決定する。このため、マスキング閾値算出部４０は、まず、周波数帯域でのマスキング閾値Ｔを算出する。

[0042] 図３（ａ）により、聴覚におけるマスキング効果について説明する。聴覚におけるマスキング効果は、ある音が他の音の存在によって聴こえにくくなる効果である。以下、聴こえを妨害する音を「マスカー」といい、聴こえが妨害される音を「マスキー」という。

マスキング効果は、周波数マスキング（同時マスキング）及び時間マスキング（継時マスキング）に大別される。周波数マスキングは、マスカーとマスキーが時間的に重なっている場合に生じるマスキングであり、時間マスキングは時間的に離れている場合に生じるマスキングである。

図３（ａ）のグラフにおいて、横軸は周波数、縦軸は信号のエネルギーである。すなわち、図３（ａ）は、ある信号に含まれるある１本のスペクトル（純音）をマスカーとしたときに、このマスカーによってマスクされるスペクトル（マスキー）の範囲及び閾値の例のグラフである。このように、信号成分の存在しないマスカーの周波数近傍についても、マスキーの閾値が上昇する。また、閾値が上昇する周波数範囲はマスカーの周波数に対して対称ではなく、マスカーに対してマスキーの周波数が高いほうが低い周波数の音よりマスクされやすい。したがって、聴覚的には、マスカーはマスカーの周波数だけではなくその両側に広がった成分を持つような状況が生じる。

[0043] 図３（ｂ）により、符号化における周波数マスキング適用の概念を示す。このグラフにおいて、横軸は周波数、縦軸は信号のエネルギーである。太い黒曲線は信号のスペクトルを表す。また、灰色の曲線はマスキング閾値を表す。ここで、図３（ｂ）において塗りつぶされている範囲が、周波数マスキングによってマスクされ知覚されない部分となる。このとき、図３（ｂ）に

おいて実際に音の知覚に寄与する部分は、信号のスペクトルを表す曲線とマスキング閾値を表す曲線に挟まれた部分となる。また、図３（ｂ）における高域のように、信号スペクトルのエネルギーがマスキング閾値より小さくなる周波数は、音の知覚に寄与しない。つまり、信号スペクトルのエネルギーからマスキング閾値を引いたエネルギーに応じたビットのみを割り当てることによっても、聴覚的には劣化が知覚されない状態で信号を伝送することが可能となる。このように、周波数領域でのマスキング効果を用いることで、伝送に必要なビット数を聴覚的な品質を保持したまま削減することが可能である。

[0044] なお、図３（ｂ）のような全帯域にわたるマスキング閾値を表す曲線は、単一のスペクトル又は雑音に関するマスキングの知見を用いて、各周波数成分に関するマスキング閾値を計算し、それらを総合することによって得られる。

[0045] ここで、この周波数帯域でのマスキング閾値Ｔの詳細な計算方法について説明する。

マスキング閾値算出部４０は、例えば、特許文献１に記載されたようなＢａｒｋスペクトルにマスキング閾値計算式（Spreading Function、以下、「ＳＦ」という。）を畳み込む。そして、マスキング閾値算出部４０は、Ｓｐｅｃｔｒａｌ　Ｆｌａｔｎｅｓｓ　ｍｅａｓｕｒｅ（ＳＦＭ）及び調整係数を用いて、Ｓｐｒｅａｄマスキング閾値T_{spread}を算出する。この上で、マスキング閾値算出部４０は、逆畳み込みにより、Ｓｐｒｅａｄマスキング閾値T_{spread}を、Ｂａｒｋスペクトルの領域に戻すことで、仮の閾値Ｔを算出する。この上で、本実施形態においては、マスキング閾値算出部４０は、仮の閾値Ｔを、各Ｂａｒｋインデックスに該当するＤＦＴスペクトルの本数で割ってから、絶対閾値と比較することで、仮の閾値Ｔが、周波数マスキングの最終的な閾値T_{final}に変換される。

[0046] より具体的に説明すると、マスキング閾値算出部４０が仮の閾値Ｔと比較する絶対閾値として、周波数ｆ（Ｈｚ）における絶対閾値の近似式T_{qf}［ｄＢ

ＳＰＬ］は、下記の式（３）により算出される。

T_{qf}＝３．６４（ｆ／１０００）$^{-0.8}$－６．５ｅｘｐ｛－０．６（ｆ／１０００－３．３）^{2}｝＋１０$^{-3}$（ｆ／１０００）^{4}＋O_{LSB}…… 式（３）

ここで、式（３）で加えられるO_{LSB}は、周波数４ｋＨｚの時の絶対閾値T_{q4000}＝ｍｉｎ（T_{qf}）が、周波数４ｋＨｚ／振幅１ｂｉｔの信号のエネルギーに一致するようなオフセット値である。

[0047] 具体的には、マスキング閾値算出部４０は、周波数マスキングのｉ番目の周波数帯域（最終帯域）における閾値T_{final}を、下記の式（４）により算出する。

[0048] [数2]

$$T_{\mathrm{final}_i} = \max\left(\frac{T_i}{k_i},\quad 10^{\frac{T_{\mathrm{q_mean}_i}}{10}}\right) \qquad \cdots\cdots \text{式（４）}$$

$$T_{\mathrm{q_mean}_i} = \frac{T_{\mathrm{q}_{bl_i}} + T_{\mathrm{q}_{bh_i}}}{2}$$

bl_i: i番目の臨界帯域の下限
bh_i: i番目の臨界帯域の上限
T_i: i番目の臨界帯域における(Bark領域での)マスキング閾値
k_i: i番目の臨界帯域に該当するFFTスペクトルの本数

[0049] この上で、マスキング閾値算出部４０は、この周波数帯域の閾値T_{final}から、聴覚の空間的マスキング効果に対応したマスキング閾値を更に算出する。この際、マスキング閾値算出部４０は、音響信号の方向情報を用いて、空間的マスキングを考慮した周波数マスキング閾値を計算する。

[0050] 図３（ｃ）により、聴覚の空間的マスキング効果に対応したマスキング閾値について説明する。

従来の音響符号化方式におけるマスキング閾値の計算では、多くの場合で

、自身のチャンネルのマスキング閾値は自身のチャンネルの信号成分のみを用いて計算している。つまり、チャンネルが複数存在する音響信号においては、対象チャンネル以外のチャンネルの信号によるマスキングを対象チャンネルのマスキングに考慮せず、各チャンネル独立にマスキング閾値を決定することとなる。

ここで、本実施形態で用いるような空間サンプリングされた音響信号は、隣接するチャンネル間での信号の相関が大きく、波形が類似した部分とそうでない部分が混在していると考えられる。したがって、マスキングの観点から考えると、空間サンプリングされた信号の符号化には、各チャンネルにおけるマスキングの情報をチャンネル間で相互に適用できる可能性がある。そこで本実施形態では、空間サンプリングされた信号の符号化のために、マスキング効果を空間領域に拡張した「空間的マスキング」を用いる。

[0051] 図３（ｃ）の概念図では、横軸は信号の空間的方向、奥行きは周波数、縦軸は信号のエネルギーを表す。マスカーの信号の裾野にある四角錐の内側の領域がこの信号によりマスクされるであろう領域を表す。図３（ｂ）の周波数マスキングと比較すると、図３（ｃ）では、方向の次元が追加されており、次元が一つ増えていることがわかる。なお、空間的方向には方位角及び仰角が含まれる。図３（ｃ）のように、空間的マスキングでは、マスキング閾値を表す曲線は３次元的になる。つまり、空間方向においてもマスキングが及び、マスクされる信号が生じる。このような空間的マスキングでは、両耳情報が相互作用する聴覚の中枢系に関わるマスキングとなる。

[0052] 図４により、空間的マスキングのマスキング閾値の計算について説明する。図４は、１からＮまでのＮ方向の信号のうち、ｉ方向の信号に対して、空間的マスキングを考慮したマスキング閾値を計算する例である。各グラフの横軸は周波数、縦軸は信号のエネルギーである。各グラフ共に、黒実線が信号スペクトルを表し、灰色実線がそれらより計算されるマスキング閾値を表す。黒の破線は、各方向の信号のマスキング閾値に重み付けを行ったものである。灰色の点線は、各方向の信号によるマスキングをすべて考慮した、ｉ

方向の信号のマスキング閾値を表す。

[0053] より具体的に説明すると、本発明者らは、後述する実施例の聴取実験の結果を踏まえ、全方位音源における空間的マスキングを考慮したマスキングモデルを作成し、下記のように計算を行った。

計算手順は次のようになる。まず、各方向の信号に関して、従来の周波数領域マスキングと同様の考え方でマスキング閾値を計算する。次に、それらの各方向のマスキング閾値Ｔを得るために、各チャンネル信号の周波数領域におけるマスキング閾値に乗ずる重みを、上述の式（１）に対応した関数$T_{spatial}$（θ，ｘ）により算出し、それぞれ重み付けする。ただし、自身すなわちｉ方向の信号のマスキング閾値に対する重み付けはゼロｄＢ、すなわち、リニアスケールでは１となるようにする。次に、重み付けされた全方向のマスキング閾値をリニアスケールで総和する。これにより、空間的マスキングを考慮したｉ方向の信号のマスキング閾値が得られる。以上の処理を、他の方向の信号についても同様に行うことで、空間的マスキングを考慮した閾値を全周の信号に対して得ることができる。

[0054] 関数$T_{spatial}$の詳細について以下に説明する。関数$T_{spatial}$は、マスカーの方位及びマスキーの方位を変数として入力したときに、マスカーの存在する方位からのマスキング閾値の減衰量をデシベルで出力する関数である。したがって、$T_{spatial}$はマスカーの存在する方位で最大値が０［ｄＢ］となるように決定する。

本実施形態においては、マスカーの方位を［ｄｅｇ．］、マスキーの方位をｘ［ｄｅｇ．］として、関数$T_{spatial}$（θ，ｘ）［ｄＢ］を、下記、式（４の２）で算出する。

[0055] $T_{spatial}(\theta, x) = \beta\{\max(f(x-\theta), \alpha f(180^\circ - x - \theta)) - 1\}$ …… 式（４の２）

ここで、α，βはスケーリング係数であり、$0 \leqq \alpha \leqq 1$，$0 \leqq \beta$である。ｍａｘは、引数内の最大値を返す関数である。ｆは、位相０°で最大値をとるような周期３６０°の任意の周期関数とする。

[0056] 本実施形態においては、この周期関数ｆ（ｘ）として、例えば、上述の式（２）と同様の三角波を用いることが可能である。このように関数ｆを定義すると、ｆ（ｘ－θ）は、マスカーの存在する方位で０ｄＢとなり、それとは正反対の方位、すなわち１８０°進んだ方位でレベルが最小となるような閾値の変化を表す。それに対して、ｆ（１８０－ｘ－θ）はマスカーの存在する方位に対して前後対称の方位で０ｄＢとなり、それとは正反対の方位、すなわち１８０°進んだ方位でレベルが最小となるような閾値の変化を示す。つまり、「マスカーの存在する方位からの閾値の減衰」及び「マスカーの存在する方位に対して前後対称となる方位からの閾値の減衰」をそれぞれ表現するように位相を合わせた関数ｆを２つ用意し、それらの最大値をとってスケーリングすることにより、「マスキーがマスカーから離れた方位にあるほど閾値が減少する現象」及び「閾値が前頭面で折り返されるような現象」の２つを同時に表現したマスキング閾値を算出可能となる。

[0057] スケーリング係数α（０≦α≦１）は、「マスカーの周波数（中心周波数）が低いほど、マスキーがマスカーに対して前後対称の方位にあるときの閾値の上昇が顕著にみられる」というマスキング効果を反映するための係数である。αは、マスカーの周波数が低いほど１に近づき、マスカーの周波数が高いほど０に近づくように決定する。そうすることで、ｆ（１８０－ｘ－θ）を、マスカーの周波数に応じてスケーリングし、閾値の前頭面での折り返し度合いを調整することが可能となる。

[0058] スケーリング係数β（０≦β）は、「マスカーが純音のときには、マスキーの方位による閾値の変化はフラットである」という知見を反映するための係数である。βはマスカーの調性がトーン性であるほど０に近づき、マスカーの調性がノイズ性であるほど値が大きくなるように決定する。そうすることで、θ及びｘが変化したときの、関数$T_{spatial}$全体としての値の振れ幅を、マスカーが純音かノイズかに応じて調整することが可能となる。

[0059] このように、本実施形態では、各チャンネル信号の周波数領域におけるマスキング閾値に乗ずる重みＴを適用する。この重みを乗じた各方向の周波数

領域マスキングの閾値を足し合わせることで、当該方向（ｘ方向）のマスキング閾値が（周波数軸上で）算出可能となる。

[0060] なお、α，βは、実施例で示したように、実際の実験により総当たりすることにより、周波数及びＳＦＭに対応した最適値を算出し、これをテーブルとして当てはめることも可能である。

[0061]（ステップＳ１０３）

次に、情報量決定部５０が、情報量決定処理を行う。

本実施形態の音響システムＸでは、空間サンプリングされた信号の方向情報を利用し、空間領域を考慮したビット割り当てを周波数領域において行う。また、空間領域を考慮したビット割り当てを行うために、マスキング効果を用いる。

このため、情報量決定部５０は、マスキング閾値算出部４０により算出されたマスキング閾値により、各チャンネル及び／又は音源オブジェクトに割り振る情報量を決定する。聴覚の空間的マスキング効果に対応したマスキング閾値を用いることで、空間領域を考慮した周波数軸上のビット割り当てを行うことが可能となる。すなわち、聴覚の空間的マスキング効果を用いることで伝送に必要な信号のビット数を聴覚的な品質を保持したまま削減可能となる。

[0062] 本実施形態において、情報量決定部５０は、聴覚のマスキング効果を積極的に利用するため、例えば、ＰＥを用いて、情報量としてビット割り当てを算出する。マスキング閾値未満の信号には人間の聴覚にとって意味のある情報は無い、すなわち量子化雑音に埋もれても良いものとして音楽信号の持つ平均情報量を計算したものがＰＥである。

このＰＥは、下記、式（５）により算出可能である。

[0063]

[数3]

$$PE(j)=\frac{1}{N}\sum_{i=1}^{25}\sum_{\omega=bl_i}^{bh_i}\left\{\log_2\left(2\left|\mathrm{round}\left(\frac{\mathrm{Re}(\mathrm{X}(\omega))}{\sqrt{6T_i/k_i}}\right)\right|+1\right)+\log_2\left(2\left|\mathrm{round}\left(\frac{\mathrm{Im}(\mathrm{X}(\omega))}{\sqrt{6T_i/k_i}}\right)\right|+1\right)\right\}\quad[\mathrm{bits/sample}]\cdots\cdots\text{式（５）}$$

N: フレーム内のサンプル数
bl_i: i番目の臨界帯域の下限
bh_i: i番目の臨界帯域の上限
T_i: i番目の臨界帯域におけるマスキング閾値
k_i: i番目の臨界帯域に該当するFFTスペクトルの本数
$\mathrm{X}(\omega)$: 各チャンネル信号の複素スペクトル

[0064] ここで、T_iは、Ｂａｒｋ領域での臨界帯域の閾値となり、$T_i/k_i=T_{final\ i}$として挿入される。

[0065]（ステップＳ１０４）

次に、符号化部６０が、符号化処理を行う。

符号化部６０は、複数のチャンネルの音響信号、及び／又は、音源オブジェクト並びに音源オブジェクトの位置情報を、それぞれ割り振られた情報量で符号化する。

符号化されたデータは、受信側の復号化装置２へと伝送される。この伝送は、例えば、ピアツーピア通信により行われる。又は、データとしてダウンロードされたり、メモリーカードや光学記録媒体として復号化装置２に読み込まれたりしてもよい。

[0066]（ステップＳ２０３）

ここで、復号化装置２の、復号化部９０が、復号化処理を行う。

復号化部９０は、符号化装置１で符号化された複数のチャンネルの音響信号、及び／又は音源オブジェクトを音声信号に復号化する。具体的には、復号化装置２がスマートフォン等の場合、符号化装置１で伝送された音響信号を、特定のコーデック等のデコーダー等で復号化する。

[0067]（ステップＳ２０４）

次に、立体音響再生部１００が、立体音響再生処理を行う。

立体音響再生部１００は、復号化部９０により復号化された音声信号を、

受聴者に対する立体音響を再生するような立体音響信号に変換する。

具体的には、立体音響再生部１００は、多チャンネルの音声信号を２チャンネルの音声信号として空間的情報を含めたまま再生する。これは、各音声信号に音源から人間の耳元までの音の伝達特性を付加し、全方向にわたって加算することにより実現可能である。つまり、立体音響再生部１００は、方向別の音信号を合成し、ヘッドフォンを用いて再生させる。このため、各音声信号の方向に対応する頭部伝達関数（ＨＲＴＦ）を畳み込み、２チャンネルの音信号に変換する。具体的には、立体音響再生部１００は、例えば、各音響信号に、各信号の方向に対応するＨＲＴＦの伝達特性を付加し、Ｌチャンネル、Ｒチャンネルそれぞれにおいて信号の総和をとって出力する。これにより、ヘッドフォンによる２チャンネルの音声信号として再生することが、収音側のチャンネル数に依存せず、手軽に再生できる。

以上により、本発明の実施の形態に係る音響信号符号化復号化処理を終了する。

[0068] 以上のように構成することで、以下のような効果を得ることができる。

近年、音響再生環境の多チャンネル化、あるいはＡＲ（拡張現実）やＶＲ（仮想現実）に於けるバイノーラル再生の普及とともに、３Ｄ音場の収音、伝送、再生、強調技術の重要性が増している。

[0069] ここで、空間サンプリングされた信号の符号化では、聴取者を取り囲む全周の音信号を対象とする必要があるため、サンプリングする方向が増えるほどチャンネル数が膨大となり、より高い合計ビットレートが必要となる。

例として、スマートフォン等を用いてインターネットを介して伝送することを考える。音楽配信サービスの１つであるＳｐｏｔｉｆｙ（登録商標）では、ストリーミング再生時のビットレートは２チャンネルのステレオで最高３２０ｋｂｐｓ程度となっている。空間サンプリングでは２チャンネルより多いチャンネル数の信号を伝送することが想定されるので、１チャンネルあたりのビットレートをより低ビットレート化する必要があった。

一方、従来、音響信号の符号化（ＭＰＥＧ等のデータ圧縮）に於いては、

聴覚のマスキング効果が利用されてきた。しかしそのマスキングは、主に周波数軸上のマスキング効果のみが用いられてきた。ＭＰＥＧ－２　ＡＡＣ、ＭＰＥＧ－４　ＡＡＣや、ＭＰ３等の音響符号化においても、多チャンネル信号の符号化においても、チャンネル毎の周波数軸における聴覚のマスキング効果が利用されてきた。

しかし、一般に多チャンネル信号によって表現される音場は、空間的に散在する複数の音源から構成される。これについて、同時刻に複数の音源が空間的に配置された際の相互のマスキング効果や聞こえについては、その作用、効果が明らかにされておらず、応用に至っていなかった。すなわち、３次元空間に配置された音源が相互にどのようなマスキング効果を与え、どのように影響を及ぼしながら聴覚に関する知覚が形成されるのかについては、何も知られていなかった。すなわち、従来のマスキング閾値の計算では、チャンネル同士の空間的な関係は考慮されていなかった。

[0070] これに対して、本発明の実施の形態に係る符号化装置１は、複数のチャンネルの音響信号、及び／又は、音源オブジェクト並びに該音源オブジェクトの位置情報を符号化する符号化装置であって、聴覚の空間的マスキング効果に対応したマスキング閾値を算出するマスキング閾値算出部４０と、マスキング閾値算出部４０により算出されたマスキング閾値により、各チャンネル及び／又は音源オブジェクトに割り振る情報量を決定する情報量決定部５０と、複数のチャンネルの音響信号、及び／又は、音源オブジェクト並びに音源オブジェクトの位置情報を、それぞれ割り振られた情報量で符号化する符号化部６０とを備えることを特徴とする。

このように構成し、複数チャンネルの音響信号又は音源オブジェクトとその位置情報を符号化する際に、聴覚の空間的なマスキング効果を勘案して各チャンネル及び音源オブジェクトに割り振るビット数を決めることで、方向情報を持った多チャンネル信号の圧縮に応用できる。これにより、チャンネル同士の空間的な関係を考慮した符号化が可能となる。

[0071] ここで、従来のマスキング閾値の計算では、チャンネル同士の空間的な関

係は考慮されていなかったため、２２．２チャンネル音響等、より臨場感を高めたチャンネル数が多い音響信号では、ビット割り当てによる圧縮が十分できず、伝送時等のビットレート（帯域）が不足するおそれがあった。

これに対して、本発明の実施の形態に係る音響信号符号化方法では、多チャンネル信号によって表現される音場は、空間的に散在する複数の音源から構成される。空間サンプリングされた信号には空間的情報が含まれるため、従来の周波数領域に加えて空間領域も考慮したビット割り当てを行うことで、より伝送ビット数を削減することも可能になる。

これにより、２２．２チャンネル等、チャンネル数が多い音響信号でも十分なビットレートでの符号化が可能な音響信号符号化方法を提供することができる。つまり、空間的に散在する複数の音源について、相互のマスキング効果に基づいてマスキング閾値を求め、その閾値に基づいたビット割り当てを行うことで、ビットレートを削減できる。本発明者らの実験によれば、従来より５～２０％ビットレートを削減可能である。

[0072] 本発明の音響システムＸは、記載の符号化装置１と、復号化装置２とを備えた音響システムであって、復号化装置２は、受聴者の向いている方向を算出する方向算出部７０と、方向算出部７０により算出された方向を符号化装置１に送信する送信部８０と、符号化装置１で符号化された複数のチャンネルの音響信号、及び／又は音源オブジェクトを音声信号に復号化する復号化部９０を備え、符号化装置１のマスキング閾値算出部４０は、マスキング閾値を、受聴者の位置と方向に対する、各チャンネル間及び／又は各音源オブジェクト間の空間的距離及び／又は方向に基づいた空間的マスキング効果に対応して算出することを特徴とする。

このように構成することで、上述の聴覚の空間的マスキング効果に対応したマスキング閾値を用いて符号化で符号化された音響信号を復号化する際に、ヘッドトラッキング等によって受聴者の向いている方向情報を算出し、音像の位置を制御する聴覚ディスプレイを実現できる。すなわち、各チャンネルの音源の位置、又は音源オブジェクトの位置と受聴者との相対的な位置関

係を、符号化装置１にフィードバックし、その位置関係に基づいて符号化を行わせ、復号化を行わせることが可能となる。

これにより、３６０°、全天球の音空間をユーザー間で手軽に収音、伝送、再生して楽しむことができる音響システムを提供できる。

[0073] 従来、３Ｄ（三次元）音場再生技術としては、音楽や放送・映画コンテンツを、ヘッドフォンや２個のフロントスピーカーでサラウンドとして楽しむバイノーラル／トランスオーラルによる聴覚ディスプレイ技術、ホームシアター向けの５．１チャンネルや７．１チャンネルサラウンド再生環境で実在するホールや劇場の音場を模擬する音場再現技術等が開発されてきた。更にスピーカーアレーによる波面合成を用いた３Ｄ音場再生技術の開発も進んでいる。このような再生方式の進化とともに、収音及びコンテンツ表現の多チャンネル化が一般化してきている。

しかしながら、３Ｄ音響の再生技術としては、頭部伝達関数と定位に関する実施形態は盛んに行われているが、空間的マスキングとの関連は検討されていなかった。

これに対して、本発明の音響システムは、復号化装置２は、復号化部９０により復号化された音声信号を、受聴者に対する立体音響を再生するような立体音響信号に変換する立体音響再生部１００を更に備えることを特徴とする。

このように構成することで、３次元空間の音場に散在する複数の音源の相互関係やマスキング効果を適用して効率的に符号化された音響信号を、空間的な音響信号の知覚に関して、頭部伝達関数（ＨＲＴＦ）と関連付けて、２チャンネルで再生できる。すなわち、人間が３Ｄ音場をどのように捉えているかに対応して符号化された音響信号を立体音響として再生することで、従来よりも現実感の高い音場を再生できる。

これは、画像において「忠実に色再現するよりも、人間が受ける「印象」を「記憶色」として再現することでよりリアル感が増す」といった効果と同様の効果と考えられる。すなわち、より現実感が高い音場再現を実現するこ

とが可能となる。

[0074] 本発明の音響信号符号化方法は、マスキング閾値は、各チャンネル間及び／又は各音源オブジェクト間の空間的距離及び／又は方向に基づいた空間的マスキング効果に対応して算出されることを特徴とする。

このように構成し、例えば、各チャンネル間及び／又は各音源オブジェクト間の空間的距離若しくは方向に基づいて算出したモデルを用いて、空間的マスキング効果に基づいた符号化が可能となる。すなわち、ヒトが３次元空間上に散在する音を聴くときに、空間的に配置された音源の空間的距離及び／又は方向に基づいた相互のマスキング効果を符号化に応用することで、より効率的な符号化を可能にし、データの伝送ビットレートを削減できる。

[0075] 本発明の音響信号符号化方法は、マスキング閾値は、チャンネル及び／又は音源オブジェクト間の空間的距離及び／又は方向が近づくほど相互に及ぼす影響が大きくなり、離れるほど相互に及ぼす影響が小さくなる空間的マスキング効果に対応して算出されることを特徴とする。

このように構成し、例えば、チャンネル及び／又は音源オブジェクト間の空間的距離あるいは方向が近づくほどチャンネル相互及び／又は音源オブジェクト相互に及ぼす影響が大きくなり、離れるほどその影響が小さくなるというモデルにより、空間的マスキング効果を算出することができる。このような空間的マスキング効果により、更に効率的な符号化を可能にし、データの伝送ビットレートを削減できる。

[0076] 本発明の音響信号符号化方法は、マスキング閾値は、受聴者からみて前後対称の位置にあるチャンネル及び／又は音源オブジェクトについては、音源オブジェクト間の空間的距離及び／又は方向についての相互に及ぼす影響の度合いを変化させる空間的マスキング効果に対応して算出されることを特徴とする。

このように構成し、受聴者からみて前後対称の位置にあるチャンネル又は音源オブジェクトについては、必ずしも音源オブジェクト間の空間的距離あるいは方向が近づくほどチャンネル相互又は音源オブジェクト相互に及ぼす

影響が大きくなり、離れるほどその影響が小さくなるというわけではないモデルにより、空間的マスキング効果を算出することができる。これにより、例えば、マスカーと前後対称の位置では空間的距離が離れるのに影響が強くなるといった空間的マスキング効果に対応して、マスキング閾値の上昇を大きく算出することが可能である。

このような空間的マスキング効果により、更に効率的な符号化を可能にし、データの伝送ビットレートを削減できる。

[0077] 本発明の音響信号符号化方法は、マスキング閾値は、受聴者からみて後方の位置にあるチャンネル及び／又は音源オブジェクトについては、前後対称の位置に該当する前方に当該チャンネル及び／又は当該オブジェクトが存在する空間的マスキング効果に対応して算出されることを特徴とする。

このように構成し、受聴者からみて後方の位置にあるチャンネル又は音源オブジェクトについては、前後対称の位置に該当する、鏡写しにした前方に当該チャンネル又は当該オブジェクトが存在する空間的マスキング効果を用いたマスキング閾値を算出することができる。すなわち、両耳を結ぶ直線を軸に、その軸より後方にある音源は、その軸を中心とする線対称の位置に該当する、軸の前方に移動するようにマスキング閾値を算出する。

このような空間的マスキング効果により、更に効率的な符号化を可能にし、データの伝送ビットレートを削減できる。

[0078] 本発明の音響信号符号化方法は、マスキング閾値は、各チャンネル及び／又は音源オブジェクトの信号が、トーン性の信号かノイズ性の信号かに対応して、各チャンネル及び／又は音源オブジェクトの信号の相互に及ぼす影響の度合いを変化させる空間的マスキング効果に対応して算出されることを特徴とする。

このように構成し、空間的マスキング効果として、各チャンネル信号又は音源オブジェクトが、トーン性の信号かノイズ性の信号かに応じて、各チャンネル信号又は音源オブジェクト信号相互に及ぼす影響の度合いを変化させるというモデルにより、マスキング閾値を算出することができる。

このように構成することで、更に効率的な符号化を可能にし、データの伝送ビットレートを削減できる。

[0079] 本発明の音響信号符号化方法は、マスキング閾値は、下記式（１）で調整される

Ｔ＝β｛ｍａｘ（ｙ１、αｙ２）－１｝

ｙ１＝ｆ（ｘ－θ）

ｙ２＝ｆ（１８０－ｘ－θ）　　……　式（１）

ただし、Ｔは前記マスキング閾値を算出するために、各チャンネル信号の周波数領域におけるマスキング閾値に乗ずる重み、θはマスカーの方位、αはマスカーの周波数で制御される定数、βはマスカーの信号がトーン性の信号かノイズ性の信号かに対応して制御される定数、ｘは前記方向又はマスキーの方位を示すことを特徴とする。

このように構成することで、上述の各モデルに対応した空間的マスキング効果を容易に計算することができる。これにより、効率的な符号化を可能にし、データの伝送ビットレートを削減できる。

[0080] 従来、ステレオ信号の各チャンネルの周波数領域におけるマスキング効果のみを考慮してＰＥを算出するのが一般的であった。

これに対して、本発明の音響信号符号化方法は、チャンネル間にまたがる空間的マスキング効果を考慮して、ＰＥにより、一サンプル当たりの平均ビット数が算出されることを特徴とする。

このように構成してマスキング閾値に対するビットの割り当てが行われると、データの伝送ビットレートを削減できる。本発明者らの実験によると、５～２５パーセント程度のビットレートを削減できることを確認している。

[0081] 本発明の音響信号復号化方法は、復号化装置２により実行される音響信号復号化方法であって、上述の音響信号符号化方法により符号化された複数のチャンネルの音響信号を復号化することを特徴とする。

このように構成し、上述の符号化装置１で符号化された音響信号を復号化することで、伝送ビットレートが低くても、高品質な音響信号を再生可能と

なる。

[0082] 〔他の実施の形態〕

なお、本発明の実施の形態においては、複数のチャンネルの音響信号の符号化として、２２．２チャンネルの符号化について言及した。

これについて、本実施形態の音響信号符号化方法は、５．１チャンネルや７．１チャンネル等の多チャンネルの音響符号化から、空間をサンプリングした３Ｄ音響符号化、ＭＰＥＧ－Ｈ　３Ｄ　ＡＵＤＩＯに代表されるオブジェクト符号化、又は、既存の２チャンネルのステレオ音響符号化にも適用可能である。

すなわち、符号化装置１は、上述の実施形態の図１に示したような、マイクロホンアレイ１０を用いて収音を行わずに、図２のステップＳ１０１にて、既に集音された多チャンネルの音声データ、音声オブジェクト等からも音声データを取得可能であるのが当然である。

[0083] さらに、上述の実施の形態では、音響システムＸが、伝送された音響信号を復号化する復号化装置２としてヘッドトラッキングが可能なヘッドフォンを用いる例について記載した。

しかしながら、本実施形態の音響信号符号化方法、及び音響復号化方法は、３次元空間的に散在する音源に働く聴覚のマスキング効果を用いることが可能な音響システムであれば、任意のものに適用可能である。たとえば、それ以外の３Ｄ音場のキャプチャー、伝送、再生システムへの適用、ＶＲ／ＡＲアプリケーションへの適用等も可能である。

[0084] 具体的な例を挙げて説明すると、上述の実施の形態では、立体音響を再生するヘッドフォン１１０として、装着可能なヘッドフォンやイヤフォン等を用いる例について説明した。

しかしながら、ヘッドフォン１１０は、実施例に示すように、据え置き型の複数個のスピーカー等であってもよいのが当然である。

[0085] さらに、上述の実施の形態では、ヘッドフォンから位置方向情報を符号化装置１へフィードバックをするように記載したものの、これをしなくてもよ

い。このように、位置方向情報のフィードバックを行わない場合、当然、当該位置方向情報を用いずに、マスキング閾値を算出することも可能である。

この場合、立体音響再生部１００は、位置方向情報に合わせて頭部伝達関数（ＨＲＴＦ）の畳み込みをアップデートしなくてもよい。

[0086] 加えて、上述の実施の形態では、復号化装置２が方向算出部７０及び送信部８０を備えている構成について説明した。

しかしながら、本実施形態の音響信号符号化方法、及び音響復号化方法は、かならずしも受聴者の向いている方向が分からなければならないということではない。このため、方向算出部７０及び送信部８０を具備しないような構成も可能である。

[0087] 上述の実施の形態では周波数マスキングを拡張した空間的マスキング効果を算出する例について記載した。

これに対して、周波数を時間に代用しても同様の空間的マスキング効果を算出することも可能である。さらに、空間的マスキング効果として、周波数、方向間でのマスキングと、時間、方向間でのマスキングとの組み合わせを用いることも可能である。

[0088] さらに、上述の実施の形態では空間的マスキング効果により、ビットレートを低く抑えたままで伝送する例について説明した。すなわち、従来の高ビットレートの音響符号化と同等の品質で、複数のチャンネルの音響信号を符号化する例について記載した。

これに対して、単に高品質での符号化を行うのみならず、重要な音を強調したり定位感をデフォルメしたりして、符号化を行うことも可能である。または、空間的マスキング効果で聴覚上、重要な箇所に割り振る情報量を増大させたり、逆に、聴覚上で重要でない箇所に割り振る情報量を更に減少させたりすることで、臨場感を強調することも可能である。

[0089] 加えて、上述の実施の形態では、情報量の割り振りとして、ビット割り当てを行う例について記載した。

しかしながら、この情報量の割り振りは、周波数帯域毎に単純にビット数

を決定（割り当てる）のではなく、エントロピー符号化やその他の符号化に対応した情報量の割り振りであってもよい。

[0090] さらに、上述の実施の形態に記載しているように、位置方向情報のフィードバックがある場合は、当該位置方向情報を用いて、効率的なマスキング閾値を算出することが可能である。

このため、位置方向情報のフィードバックの有無により、配信（伝送）のビットレートを変更するように構成することが可能である。すなわち、符号化装置１に対して、位置方向情報をフィードバックしてくる復号化装置２は、位置方向情報をフィードバックしてこない復号化装置２よりも低いビットレートでデータを伝送することが可能である。

このように構成することで、より廉価にコンテンツを提供するサービスを実現することが可能となる。

[0091] 次に図面に基づき本発明を実施例によりさらに説明するが、以下の具体例は本発明を限定するものではない。

実施例

[0092] （空間的マスキングを考慮したマスキングモデルの実験）

（実験方法）

図５、図６により、マスカー存在下でのマスキーの各周波数における閾値を、マスキーの各方位に関して測定する実験について説明する。

図５は、測定システムを示す構成図である。ここでは、被験者の正面を０°とし、反時計方向を正とする。そして、被験者の正面にＰＣ（Personal Computer）が配置される。被験者は椅子に座り、スピーカで提示された刺激音を両耳で聴取する。スピーカは、被験者から１．５ｍ離れた位置に、被験者を中心として全周を取り囲むように、４５°間隔で８か所に配置される。なお、実験系の出力における音圧レベル［ｄＢＳＰＬ］の校正は、騒音計（リオンＮＡ－２７）を用いて計測することにより行った。

実験方法を以下に記す。最初に、実験で使用する音源を被験者に把握させるために、各音源を個別に提示するデモを行う。次に、測定を開始する。測

定中、マスカーは常時提示される。マスキーは継続時間０．７秒で提示され、０．７秒の無音をはさんだ後に提示が繰り返される。被験者は回答画面を見ながら、マスキーの各周波数、各音圧レベルに対し、マスキーが３回提示される間に、「マスカー音に変化を感じたかどうか」をＰＣに入力する。この際、被験者には頭部を動かさずに視線のみを移動させて回答を入力するよう指示を与える。ここで、「マスカー音に変化を感じた」とは、マスキーが知覚されたときだけでなく、マスカーでもマスキーでもない音が知覚された場合も含むこととする。例えば、周波数が少し異なる２つの純音が同時に提示されたとき、音波の干渉により２音の周波数の差に等しい周波数の音が知覚される「うなり」が挙げられる。そのような音が知覚された場合も、「マスカーに変化を感じた」場合に含む。

なお、実験方法に慣れさせるために、実験結果に反映しないテスト測定を初めに数回、行った。

[0093] 図６に、本実験における閾値探索方法の説明図を示す。本実験における閾値の探索方法は適応法に準じた方法で行う。適応法とは、被験者の応答に応じて実験者が刺激の物理パラメータ値を調整し、閾値を決定する方法のことである。

図６において、横軸はマスキーのセット数、縦軸はマスキーの音圧レベルである。マスキーのセット数「１セット」とは、マスキーが３回提示される間のことを指し、これを音源提示の単位とする。

まず、マスキーの周波数をｆ１に固定し、音圧レベルＳＰＬｍａｘで聴取者に提示する。続いて、音圧レベルをＳＰＬｍｉｎに変更して聴取者に提示する。ＳＰＬｍａｘは音圧レベルの測定範囲における最大値、ＳＰＬｍｉｎは音圧レベルの測定範囲における最小値を指す。ここで、被験者が音圧レベルＳＰＬｍａｘのマスキーを検知できなかった場合にはＳＰＬｍａｘを閾値とみなし、音圧レベルＳＰＬｍｉｎのマスキーを検知できた場合にはＳＰＬｍｉｎを閾値とみなす。このとき、実際の閾値は測定範囲外に存在すると考えられる。以上のようにみなされる例として、図６における周波数ｆ２のマ

スキーの閾値が挙げられる。図６では、周波数ｆ２のマスキーは音圧レベルＳＰＬｍｉｎでも検知されなかったことを示している。このように、被験者が回答しなければならない音圧レベルのセット数は、被験者の応答によって変化する。マスキーが音圧レベルＳＰＬｍｉｎで提示された後は、被験者の回答に応じて閾値を２分探索的に探索する。すなわち、これまでの測定で検知できたマスキーの音圧レベルの最小値と、検知できなかったマスキーの音圧レベルの最大値の中間になるような値を、次の音圧レベルの値としてセットする。このような探索を続けると、最終的にセットできる音圧レベルが１つだけ残る。最終的に残った音圧レベルを周波数ｆ１のマスキーの閾値とする。

以上のような探索を、図６のように周波数をｆ１、ｆ２、ｆ３、……の順に連続的に変化させて調査する。本実験においては、低周波数側から順にマスキーの閾値を調査する。

[0094] 図７に、被験者に提示する回答画面を示す。マスカーが１音源のときの回答画面は図７（ａ）であり、マスカーが２音源のときの回答画面は図７（ｂ）である。画面には、マスカーの方位、マスカーの音圧レベル、マスキーの方位、マスキーの周波数、マスキー再生中に点灯するランプ、マスキーの再生回数を示すカウンタ、マスキーの検知の有無を入力するボタンがそれぞれ表示される。被験者は、各音源がどの方向からどのような大きさでいつ提示されるのかが知覚可能である。マスキーの周波数を表示する理由は、測定がマスカーの周波数（マスカーの種類）を連続的に変化させながら調査するものであるので、被験者が現在どのマスキーに関する回答を入力しているのかを明確にし、回答の混乱を防ぐためである。被験者は自ら、マスキーの検知の有無を入力するボタンをオンにすることで「マスキーを検知した」ことをＰＣへ知らせ、またボタンをオフにすることで「マスキーが検知できなかった」ことをＰＣへ知らせる。なお、マスキーの再生回数を示すカウンタの初期値は０であり、マスキーの再生回数に応じて、０、１、２、３、０ ……と変化する。０がカウントされると、回答がリセットすなわちマスキーの検

知の有無を入力するボタンがオフになり、マスキーは次の音圧レベル又は周波数に移行する。被験者は、このカウンタが１、２、３を表示している間に検知の有無を入力しなければならない。

なお、聴取実験の回答用プログラムは、Ｃｙｃｌｉｎｇ ’７４社のＭａｘ ｖｅｒ．７にてコーディングを行っている。それ以外のプログラムについては、ＭａｔｈＷｏｒｋｓ 社のＭＡＴＬＡＢ ｖｅｒ．Ｒ２０１８ａにてコーディングを行っている。

[0095] （マスカーの一覧）

実験で使用するマスカーの一覧を下記の表１に示す。

[0096] [表1]

使用するマスカー

名前	音源信号
マスカー A (maskerA)	中心周波数 400 Hz，帯域幅 100 Hz の帯域雑音
マスカー B (maskerB)	中心周波数 1000 Hz，帯域幅 150 Hz の帯域雑音
マスカー C (maskerC)	周波数 400 Hz の純音
マスカー D (maskerD)	周波数 1000 Hz の純音

[0097] マスカーには、周波数（中心周波数）を４００Ｈｚ又は１０００Ｈｚとした帯域雑音及び純音を用意した。以降では、これらのマスカーを、マスカーＡ（ｍａｓｋｅｒ Ａ）～マスカーＤ（ｍａｓｋｅｒ Ｄ）までの名前で記述することとする。なお、帯域雑音の帯域幅は、臨界帯域の帯域幅に概ね合致するように決定した。ある純音のマスクに寄与する雑音成分は、その純音を中心周波数とする帯域雑音における、ある帯域幅の成分に限られるということが知られている。臨界帯域とは、そのような純音のマスクに寄与する帯域のことである。

[0098] （実験条件）

実験条件としては、マスカーの数を１個とした場合及び２個とした場合の２種類について行った。いずれも無響室で実験を行い、音源信号のサンプリング周波数は４８ｋＨｚとした。

まず、配置するマスカーの数が１個のときの条件を下記の表２に示す。

[0099] [表2]

実験条件 (マスカーを 1 音源とするとき)

実験場所	無響室
被験者	2 名 (被験者 a，被験者 b)
サンプリング周波数 [kHz]	48
配置するマスカーの数	1
配置するマスカー	マスカー A または マスカー B または マスカー C または マスカー D
マスカーの音圧レベル [dBSPL]	60 または 80
マスカーの方位 [deg.]	0 または 45 または 90 または 135
配置するマスキーの数	1
配置するマスキー	純音
マスキーの周波数 [Hz]	マスカーの周波数または中心周波数が 400 Hz のとき: 100 200 300 340 370 390 397 403 410 430 460 500 600 700 797 803 900 1000 1197 1203 1400 1597 2000 2400 2800 3200 4000 5000 6000 マスカーの周波数または中心周波数が 1000 Hz のとき: 100 200 300 400 500 600 700 800 900 940 970 990 997 1003 1010 1030 1060 1100 1200 1400 1600 1997 2003 2400 2800 3200 3997 5000 6000
マスキーの音圧レベル [dBSPL]	マスカーの音圧レベルが 60 dBSPL のとき: 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60 マスカーの音圧レベルが 80 dBSPL のとき: 20 23 26 29 32 35 38 41 44 47 50 53 56 59 62 65 68 71 74 77 80
マスキーの方位 [deg.]	0 または 45 または 90 または 135 または 180 または 225 または 270 または 315

[0100] 被験者は、健聴な２０代の男性２名（被験者ａ、被験者ｂ）である。マスカーには、上述のマスカーＡ～マスカーＤまでの音源のうちのいずれか１つを用いた。マスカーの音圧レベルは、６０ｄＢＳＰＬ及び８０ｄＢＳＰＬの２通りを用いた。マスカーの方位は、０°、４５°、９０°、１３５°の４つの方位のうちのいずれか１方位とした。すなわち、マスカーの方位は左耳側の４方位のみ対象とした。上記のようにマスカーの方位を４方位用意して実験を行うと、被検者に関する半周分の閾値のデータが得られることとなる。人間の頭部形状が左右対称であると仮定すれば、閾値は正中面で対称になると考えられるので、本実験で得られない残り半周分の閾値のデータは本実験で得られたデータと対称の結果となる。

マスキーは純音１音源を用い、その周波数及び音圧レベルは以下の通りで

ある。具体的には、マスキーの周波数は、マスカーの周波数（中心周波数）に近い周波数では密になるように決定した。なお、マスカーが純音の場合、マスキーの周波数がマスカーの周波数と完全に一致するとき（４００Ｈｚ、１０００Ｈｚ）には、あらゆる音圧レベルにおいてマスキーが知覚できないと考えられるので、そのような周波数は測定対象から外した。マスキーの音圧レベルは取りうる値を３ｄＢおきとし、その最大レベルはマスカーの音圧レベル、最小レベルは２０ｄＢＳＰＬ又は１８ｄＢＳＰＬとした。最大レベルは、マスキーの音圧レベルがマスカーの音圧レベルより大きいときには完全にマスキーを知覚できるという予想のもとに決定した。最小レベルは、実験場所である無響室内の暗騒音レベルを考慮し、測定範囲が概ね暗騒音レベルより１５ｄＢ小さいところまでとなるように決定した。マスキーの方位は、４５°又は３１５°とした。マスキーの方位が４５°のときには、マスカーとマスキーの方位が一致するため、従来から検討されてきた周波数マスキングの閾値が結果として得られることとなる。対してマスキーの方位が３１５°のときには、マスカーとマスキーが互いに異なる方位に存在することとなるため、ステレオのチャンネル間でのマスキングすなわち空間的なマスキングの閾値が結果として得られることとなる。

マスキーの方位は、０°から４５°おきに３１５°までの８方位のうちのいずれか１方位とした。

[0101] 次に、配置するマスカーの数が２個のときの条件を下記の表３に示す。

[0102]

[表3]

実験条件 (マスカーを 2 音源とするとき)

実験場所	無響室
被験者	1 名 (被験者 a)
サンプリング周波数 [kHz]	48
配置するマスカー	マスカー A および マスカー B
配置するマスカーの数	2
マスカーの音圧レベル [dBSPL]	60 または 80
マスカーの方位 [deg.]	マスカー A: 45, マスカー B: 315
配置するマスキーの数	1
配置するマスキー	純音
マスキーの周波数 [Hz]	100 200 300 340 370 390 397 400 403 410 430 460 500 600 700 797 800 803 900 940 970 990 997 1000 1003 1010 1030 1060 1100 1197 1200 1203 1400 1597 1600 1997 2000 2003 2400 2800 3200 3997 4000 5000 6000
マスキーの音圧レベル [dBSPL]	マスカーの音圧レベルが 60 dBSPL のとき: 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60 63 66 69 マスカーの音圧レベルが 80 dBSPL のとき: 20 23 26 29 32 35 38 41 44 47 50 53 56 59 62 65 68 71 74 77 80 83 86 89
マスキーの方位 [deg.]	225

[0103] 被験者は、被験者ａのみである。マスカーは、マスカーＡを方位４５°に、マスカーＢを方位３１５°にそれぞれ配置した。マスキーは純音１音源を用いた。マスキーの周波数は、マスカーの周波数（中心周波数）が４００Ｈｚのときの条件及び１０００Ｈｚのときの条件を合わせたものを用いた。なお、配置するマスカー（マスカーＡ、マスカーＢ）がいずれもバンドノイズであるため、マスキーの周波数がマスカーの中心周波数と完全に一致するとき（４００Ｈｚ、１０００Ｈｚ）においても、純音とは異なり、ある音圧レベル以上ではマスキーを知覚できるようになると考えられる。したがって、４００Ｈｚ及び１０００Ｈｚも測定対象に加えた。また、マスキーの音圧レベルの最大値は、表２よりも９ｄＢ大きくとった。これは、マスカーが２音源存在することにより、聴取する音の音圧レベルが最大で６ｄＢほど上昇することを考慮したものである。

マスキーの方位は２２５°とした。

[0104] （マスキング閾値の計算）

（実験結果と考察）

図８～図１１により、被験者ａに関する実験結果について説明する。

[0105] 上述の式（５）に記載したα，βを、下記の表４に示す値の範囲で探索した。

[0106] [表4]

α, β の総当たりの範囲

パラメータ	総当たりの範囲
α	$0,\ 0.01,\ 0.02,\ \cdots,\ 1$
β	$0,\ 0.01,\ 0.02,\ \cdots,\ 20$

[0107] 本実施例では、α，βの最適値は次のように算出した。まず、あるα，βの値における$T_{spatial}$と、実験結果として得られたマスキーの各方位における閾値の最大値との間の平均二乗誤差（Mean Squared Error、ＭＳＥ）を、マスカーの種類（マスカーＡ～マスカーＤ）、方位、音圧レベルのすべての組み合わせに対して計算する。次に、計算された平均二乗誤差を、マスカーの種類ごとに総和をとる。以上の操作を、α，βの値を変化させて繰り返し行い、平均二乗誤差のマスカーの種類ごとの総和が最小になったときの、α，βの組を、α，βの最適値とする。

ここで、ｊ番目のマスカーの方位における平均二乗誤差ＭＳＥ（ｊ）は、下記の式（６）で算出する。

[0108] [数4]

$$MSE(j) = \frac{1}{N}\sum_{i=1}^{N}\left\{\left(T_{\mathrm{spatial}}(i) + L_{\mathrm{masker_azimuth}}\right) - T_{\mathrm{measured}}(i)\right\}^2 \cdots\cdots \text{式（６）}$$

[0109] ここで、式（６）において、$T_{spatial}$（ｉ）はｉ番目のマスキーの方位［ｄｅｇ。］における関数$T_{spatial}$の出力値、$T_{measured}$（ｉ）はｉ番目のマスキーの方位［ｄｅｇ。］におけるマスキーの閾値の実験により得られた実測値を表す。$L_{masker\ azimuth}$はマスカーの存在する方位におけるマスキーの閾値［ｄＢＳ

ＰＬ］を表す。これは、$T_{spatial}$がマスカーの存在する方位からの閾値の減衰量を表すものであるため、$T_{spatial}$と$T_{measured}$との間のオフセットを調整する役割をもつ。Ｎは$T_{spatial}$及び$T_{measured}$のエントリー数（マスキーの方位の総数）である。本計算ではマスキーの方位の刻みを０°から３６０°までの１°刻みとしたため、Ｎ＝３６１である。ただし、$T_{measured}$はマスキーの方位の刻みが実測値として４５°刻みであるため、１°刻みとしたときに欠損する部分は線形補間を行うことにより値を推定した。

総当たりの結果、α，βの最適値がマスカーＡ～マスカーＤについて、下記の表５のように得られた。

[0110] [表5]

総当たりにより得られたα, βの最適値

マスカーの種類	αの最適値	βの最適値
マスカー A	0.40	11.96
マスカー B	0.28	9.24
マスカー C	0.52	1.12
マスカー D	0.30	5.82

[0111] 図８～図１１に、表５の値を用いて$T_{spatial}$をマスキーの閾値の実測値にフィッティングさせたものをそれぞれ示す。各図の左上のグラフはマスカーＡに関する結果、右上のグラフはマスカーＢに関する結果、左下のグラフはマスカーＣに関する結果、右下のグラフはマスカーＤに関する結果である。

各グラフの横軸はマスキーの方位、縦軸は音圧レベルである。マスカーの方位に該当する方位を縦の点線で示している。黒の実線はマスカーの音圧レベルが８０ｄＢＳＰＬのときのマスキーの閾値の実測値、灰色の実線はマスカーの音圧レベルが６０ｄＢＳＰＬのときのマスキーの閾値の実測値をそれぞれ表している。これに対して、赤の破線は関数$T_{spatial}$を用いて赤の実線にフィッティングさせたもの、灰色の破線は関数$T_{spatial}$を用いて灰色の実線に

フィッティングさせたものをそれぞれ表している。

なお、各破線は関数$T_{spatial}$の出力にオフセット$L_{masker\ azimuth}$を加えたものである。

図８～図１１によれば、各グラフとも概ね実測値にフィットしていることがわかる。ただし、例えば図８の左上のグラフや図９の左上のグラフなどのように、マスカーＡ、マスカーＢのような帯域雑音の場合におけるマスカーとは前後対称の方位での閾値の上昇に関してみると、破線が実線にうまくフィットしていない部分が見受けられる。この理由は、マスカーが帯域雑音でマスカーの方位が９０°のときには、閾値の方位による変化が比較的小さく、平均二乗誤差の総和を最小にしようとしたときに影響してαの値が小さくなるように働いたためであると考えられる。上記の部分をうまくフィットさせるためには、マスカーの方位が９０°のときの実測値とモデル関数との間の誤差が大きくても構わない場合には、αの値をより大きく設定すれば良い。

また、本実施例では総当たりにより、α，βの値を求めたが、βの値に関しては、マスカーの調性（トーン性、ノイズ性）を判別するような指標をベースに決定することができる。マスカーの調性を判別するような指標としては、例えば自己相関やＳｐｅｃｔｒａｌ　Ｆｌａｔｎｅｓｓ　Ｍｅａｓｕｒｅ（ＳＦＭ）等がある。これらの指標を用いることで、βをパラメトリックに決定しフィッティングすることが可能となる。

[0112] （まとめ）

本実施例では、空間的マスキングを確認するために基礎的な聴取実験を行うとともに、実験により得られた知見を反映し、空間的マスキングを考慮したマスキング閾値計算法及びモデル化をすることが可能となった。

まず聴取実験において、マスカーとマスキーを異なる方位に存在する場合でもマスカーの周波数近傍での閾値の上昇がみられたことから、空間的マスキングの存在を確認した。

マスキング閾値はマスカーの方位とマスキーの方位によって変化し、基本

的にはマスキーの方位がマスカーの方位から離れるほど閾値が低下する。２チャンネルステレオ環境に関しては、自身のチャンネルの信号が自身のチャンネルに及ぼすマスキングの閾値に１５ｄＢの重みを付加したものを、自身のチャンネルの信号が他方のチャンネルの信号に及ぼすマスキングの閾値として用いてもよい。全方位に関しては、マスカーが帯域雑音のときは、マスカーに対して前後対称の方位でその周囲の方向よりマスキーの閾値の上昇がみられ、それはマスカーの中心周波数が低いほど顕著である。また、マスカーが純音のときは、マスキーの方位による閾値の変化はフラットである。

さらに、各マスカーが単独で存在するときの、マスカーと同一の方位の信号のマスキング閾値とそれ以外の方位の信号のマスキングの閾値とのリニアスケールでの和を、自身の方位の信号に加えそれ以外の方位の信号も考慮したマスキング閾値として用いても差し支えない。

[0113] 以下で、これらの結果をまとめると：

マスカーが０°のときは、マスキーの位置が０°のものが、もっとも閾値が高い。４５°、９０°と、マスキー位置がマスカーから離れるほど、閾値は下がった。しかし、１３５°から上昇を始め、１８０°では０°の場合とほぼ同程度まで、閾値が上昇した。すなわち、マスカーによるマスキング閾値の値が、受聴者の前後でほぼ対称の関係となっていた。

マスカーが４５°のときは、マスキー位置が４５°のときが、もっとも閾値が高くなった。９０°では、閾値が下がった。１３５°で更に下がると思われたが、予想に反し、閾値が上がり、４５°の時の閾値に近づいた。１８０°では閾値は下がり、２２５°では更に下がった。これは、マスカーが０°のときと同様に、マスキング閾値は、受聴者の前後で、ほぼ対称の関係となっている。すなわち、９０°～２７０°を結ぶ線を中心に線対称であった。

マスカーが９０°、マスカー１３５°でも、同様の傾向であった。

[0114] 以上のような知見から、空間的マスキングを考慮したマスキング閾値計算法を次のように提案した：２チャンネルのステレオ環境では、自身のチャン

ネルのマスキング閾値と、他方のチャンネルのマスキング閾値に、－１５ｄＢ重み付けしたものをリニアスケールで和をとる。全方位に関しては、周期３６０°の任意の周期関数と、その周期関数を９０°及び２７０°で線対称になるように位相シフトしたものを利用して、マスキーの閾値のピークの方位による変化をモデル化する。そのモデル化した関数を用いて、各チャンネルのマスキング閾値に重み付けをしてからリニアスケールで総和をとる。

すなわち、上述の式（１）により、マスキング閾値を計算可能となる。これに基づいてマスキング閾値を計算することで、信号の伝送に必要なビット数を削減することができる。

[0115] なお、上記実施の形態の構成及び動作は例であって、本発明の趣旨を逸脱しない範囲で適宜変更して実行することができることは言うまでもない。

産業上の利用可能性

[0116] 本発明の生物配列分析方法は、聴覚の空間的マスキング効果を利用することで、従来よりもビットレートを抑えた音響信号符号化方法を提供することができ、産業上に利用することができる。

符号の説明

[0117] １　符号化装置

２　復号化装置

１０　マイクロホンアレイ

２０　集音部

３０　周波数領域変換部

４０　マスキング閾値算出部

５０　情報量決定部

６０　符号化部

７０　方向算出部

８０　送信部

９０　復号化部

１００　立体音響再生部

１１０　ヘッドフォン

Ｘ　音響システム

請求の範囲

[請求項1] 符号化装置により実行される、複数のチャンネルの音響信号を符号化する音響信号符号化方法であって、

聴覚の空間的マスキング効果に対応したマスキング閾値を算出し、

算出された前記マスキング閾値により、各前記チャンネルに割り振る情報量を決定し、

複数の前記チャンネルの音響信号を、それぞれ割り振られた前記情報量で符号化する

ことを特徴とする音響信号符号化方法。

[請求項2] 符号化装置により実行される、音源オブジェクト及び該音源オブジェクトの位置情報を符号化する音響信号符号化方法であって、

聴覚の空間的マスキング効果に対応したマスキング閾値を算出し、

算出された前記マスキング閾値により、前記音源オブジェクトに割り振る情報量を決定し、

前記音源オブジェクト及び前記音源オブジェクトの位置情報を、割り振られた前記情報量で符号化する

ことを特徴とする音響信号符号化方法。

[請求項3] 前記マスキング閾値は、

各前記チャンネル間及び／又は各前記音源オブジェクト間の空間的距離及び／又は方向に基づいた前記空間的マスキング効果に対応して算出される

ことを特徴とする請求項１又は２に記載の音響信号符号化方法。

[請求項4] 前記マスキング閾値は、

前記チャンネル及び／又は前記音源オブジェクト間の空間的距離及び／又は方向が近づくほど相互に及ぼす影響が大きくなり、離れるほど相互に及ぼす影響が小さくなる前記空間的マスキング効果に対応して算出される

ことを特徴とする請求項３に記載の音響信号符号化方法。

[請求項5] 前記マスキング閾値は、

受聴者からみて前後対称の位置にある前記チャンネル及び／又は前記音源オブジェクトについては、前記音源オブジェクト間の空間的距離及び／又は方向についての相互に及ぼす影響の度合いを変化させる前記空間的マスキング効果に対応して算出される

ことを特徴とする請求項３又は４に記載の音響信号符号化方法。

[請求項6] 前記マスキング閾値は、

前記受聴者からみて後方の位置にある前記チャンネル及び／又は前記音源オブジェクトについては、前後対称の位置に該当する前方に当該チャンネル及び／又は当該オブジェクトが存在する前記空間的マスキング効果に対応して算出される

ことを特徴とする請求項３乃至５のいずれか１項に記載の音響信号符号化方法。

[請求項7] 前記マスキング閾値は、

各前記チャンネル及び／又は前記音源オブジェクトの信号が、トーン性の信号かノイズ性の信号かに対応して、各前記チャンネル及び／又は音源オブジェクトの信号の相互に及ぼす影響の度合いを変化させる前記空間的マスキング効果に対応して算出される

ことを特徴とする請求項３乃至６のいずれか１項に記載の音響信号符号化方法。

[請求項8] 前記マスキング閾値は、

下記式（１）で調整される

$T=\beta\{\max(y1,\alpha y2)-1\}$

$y1=f(x-\theta)$

$y2=f(180-x-\theta)$ …… 式（１）

ただし、Ｔは前記マスキング閾値を算出するために、各チャンネル信号の周波数領域におけるマスキング閾値に乗ずる重み、θはマスカーの方位、αはマスカーの周波数で制御される定数、βはマスカーの

信号がトーン性の信号かノイズ性の信号かに対応して制御される定数、ｘは前記方向又はマスキーの方位を示す

ことを特徴とする請求項７に記載の音響信号符号化方法。

[請求項9] Ｐｅｒｃｅｐｔｕａｌ Ｅｎｔｒｏｐｙ（ＰＥ）により、一サンプル当たりの平均ビット数が算出される

ことを特徴とする請求項１乃至８のいずれか１項に記載の音響信号符号化方法。

[請求項10] 復号化装置により実行される音響信号復号化方法であって、

請求項１乃至９に記載の音響信号符号化方法により符号化された複数の前記チャンネルの音響信号を復号化する

ことを特徴とする音響信号復号化方法。

[請求項11] 符号化装置により実行される、複数のチャンネルの音響信号を符号化するプログラムであって、前記符号化装置に、

聴覚の空間的マスキング効果に対応したマスキング閾値を算出させ、

算出された前記マスキング閾値により、各前記チャンネルに割り振る情報量を決定させ、

複数の前記チャンネルの音響信号を、それぞれ割り振られた前記情報量で符号化させる

ことを特徴とするプログラム。

[請求項12] 複数のチャンネルの音響信号、及び／又は、音源オブジェクト並びに該音源オブジェクトの位置情報を符号化する符号化装置であって、

聴覚の空間的マスキング効果に対応したマスキング閾値を算出するマスキング閾値算出部と、

前記マスキング閾値算出部により算出された前記マスキング閾値により、各前記チャンネル及び／又は前記音源オブジェクトに割り振る情報量を決定する情報量決定部と、

複数の前記チャンネルの音響信号、及び／又は、前記音源オブジェ

クト並びに前記音源オブジェクトの位置情報を、それぞれ割り振られた前記情報量で符号化する符号化部とを備える

ことを特徴とする符号化装置。

[請求項13] 請求項１２に記載の符号化装置と、復号化装置とを備えた音響システムであって、

前記復号化装置は、

前記符号化装置で符号化された複数の前記チャンネルの音響信号、及び／又は前記音源オブジェクトを音声信号に復号化する復号化部を備える

ことを特徴とする音響システム。

[請求項14] 請求項１２に記載の符号化装置と、復号化装置とを備えた音響システムであって、

前記復号化装置は、

受聴者の向いている方向を算出する方向算出部と、

前記方向算出部により算出された前記方向を前記符号化装置に送信する送信部と、

前記符号化装置で符号化された複数の前記チャンネルの音響信号、及び／又は前記音源オブジェクトを音声信号に復号化する復号化部を備え、

前記符号化装置の前記マスキング閾値算出部は、

前記マスキング閾値を、前記受聴者の位置と前記方向に対する、各前記チャンネル間及び／又は各前記音源オブジェクト間の空間的距離及び／又は方向に基づいた前記空間的マスキング効果に対応して算出する

ことを特徴とする音響システム。

[請求項15] 前記復号化装置は、

前記復号化部により復号化された音声信号を、前記受聴者に対する立体音響を再生するような立体音響信号に変換する立体音響再生部を

更に備える

ことを特徴とする請求項１３又は１４に記載の音響システム。

[請求項16] 聴覚の空間的マスキング効果に対応したマスキング閾値により、各チャンネル及び／又は音源オブジェクトに割り振る情報量が決定され、複数の前記チャンネルの音響信号、及び／又は、前記音源オブジェクト並びに前記音源オブジェクトの位置情報を、それぞれ割り振られた前記情報量で符号化された信号を取得する信号取得部と、

前記信号取得部により取得された信号から、符号化された複数の前記チャンネルの音響信号、及び／又は前記音源オブジェクトを音声信号に復号化する復号化部とを備える

ことを特徴とする復号化装置。

[請求項17] 受聴者の向いている方向を算出する方向算出部と、

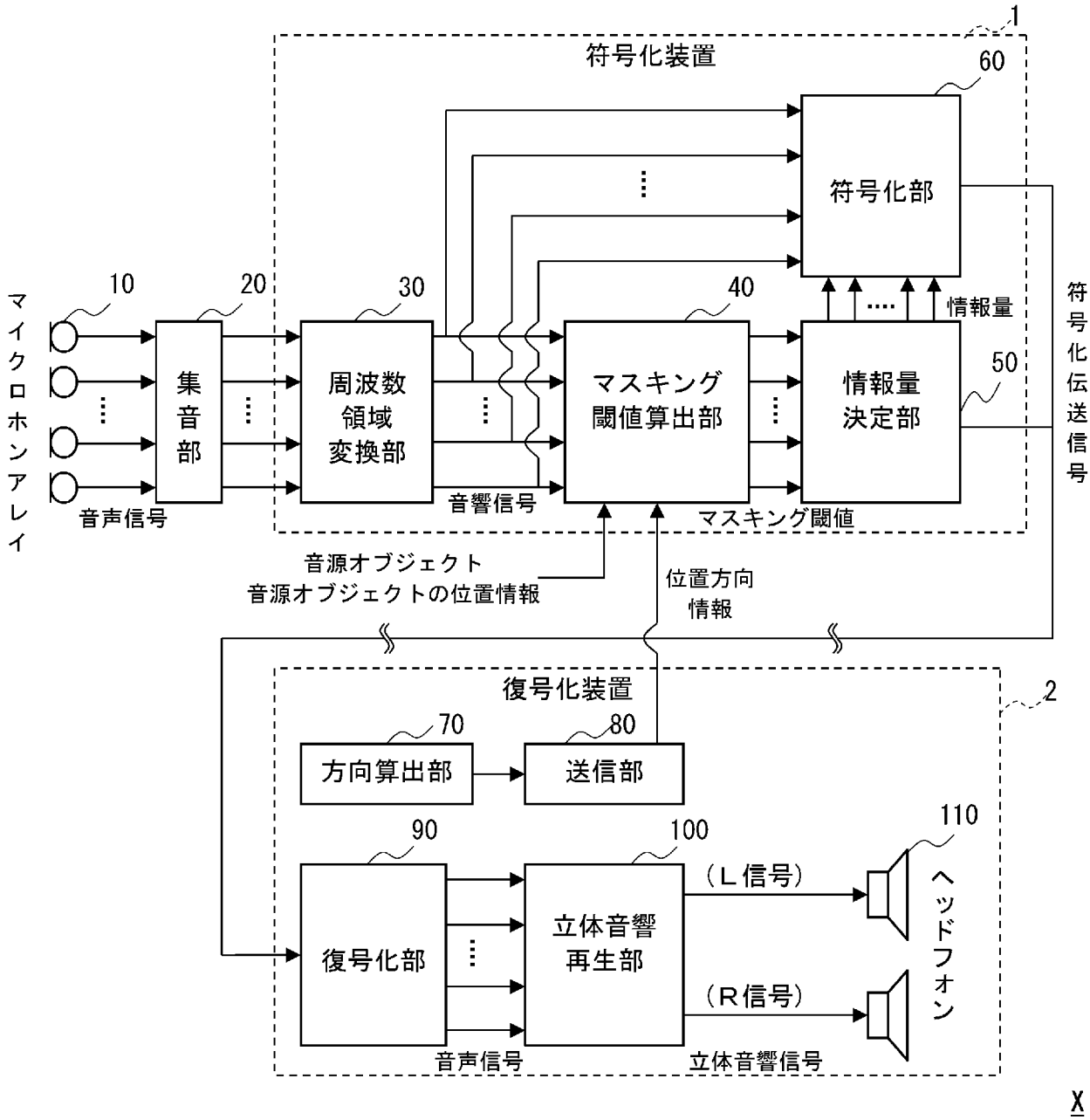
前記方向算出部により算出された前記方向を前記符号化装置に送信する送信部とを更に備える

ことを特徴とする請求項１６に記載の復号化装置。

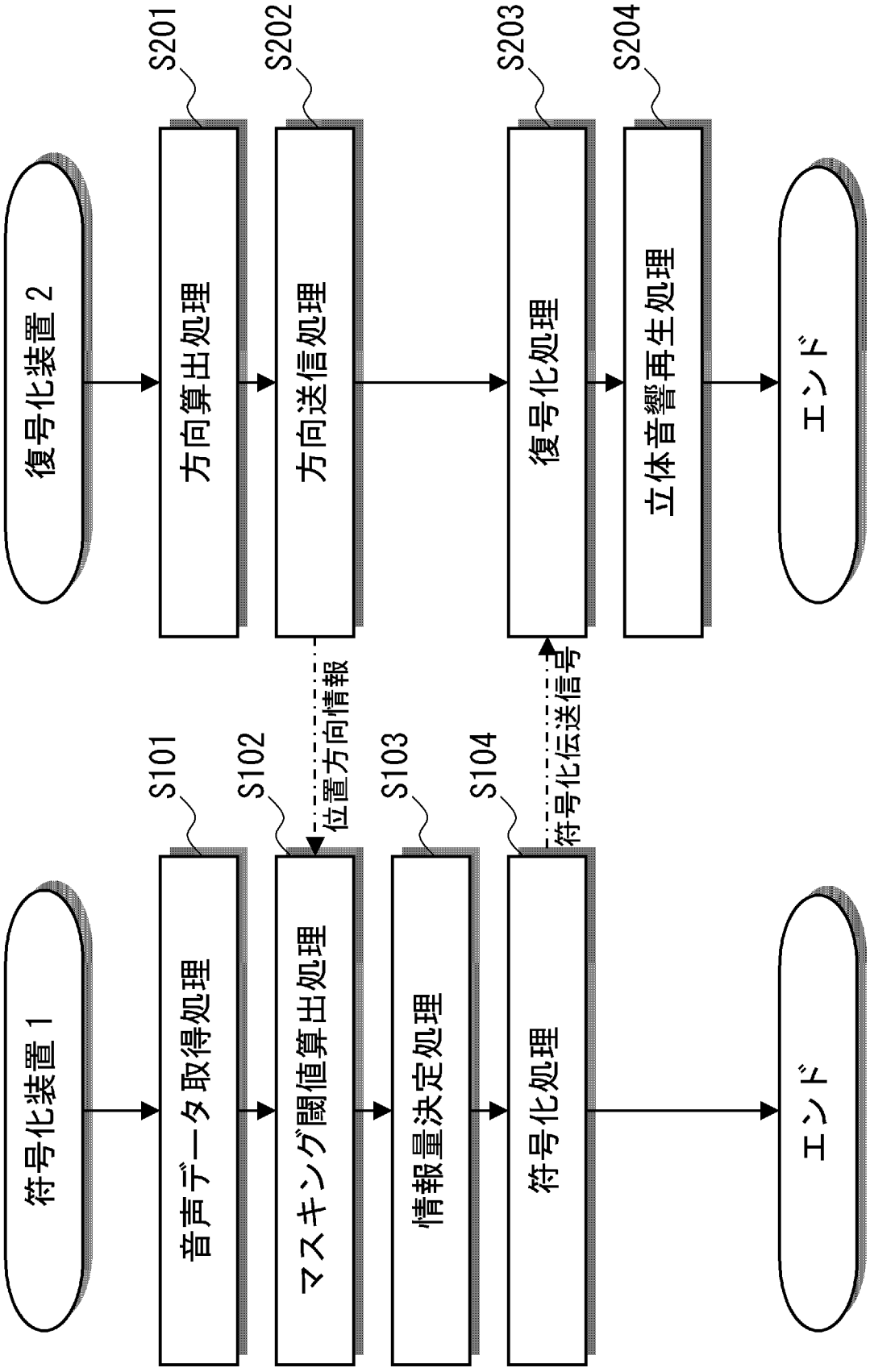
[請求項18] 前記復号化部により復号化された音声信号を、前記受聴者に対する立体音響を再生するような立体音響信号に変換する立体音響再生部を更に備える

ことを特徴とする請求項１６又は１７に記載の復号化装置。

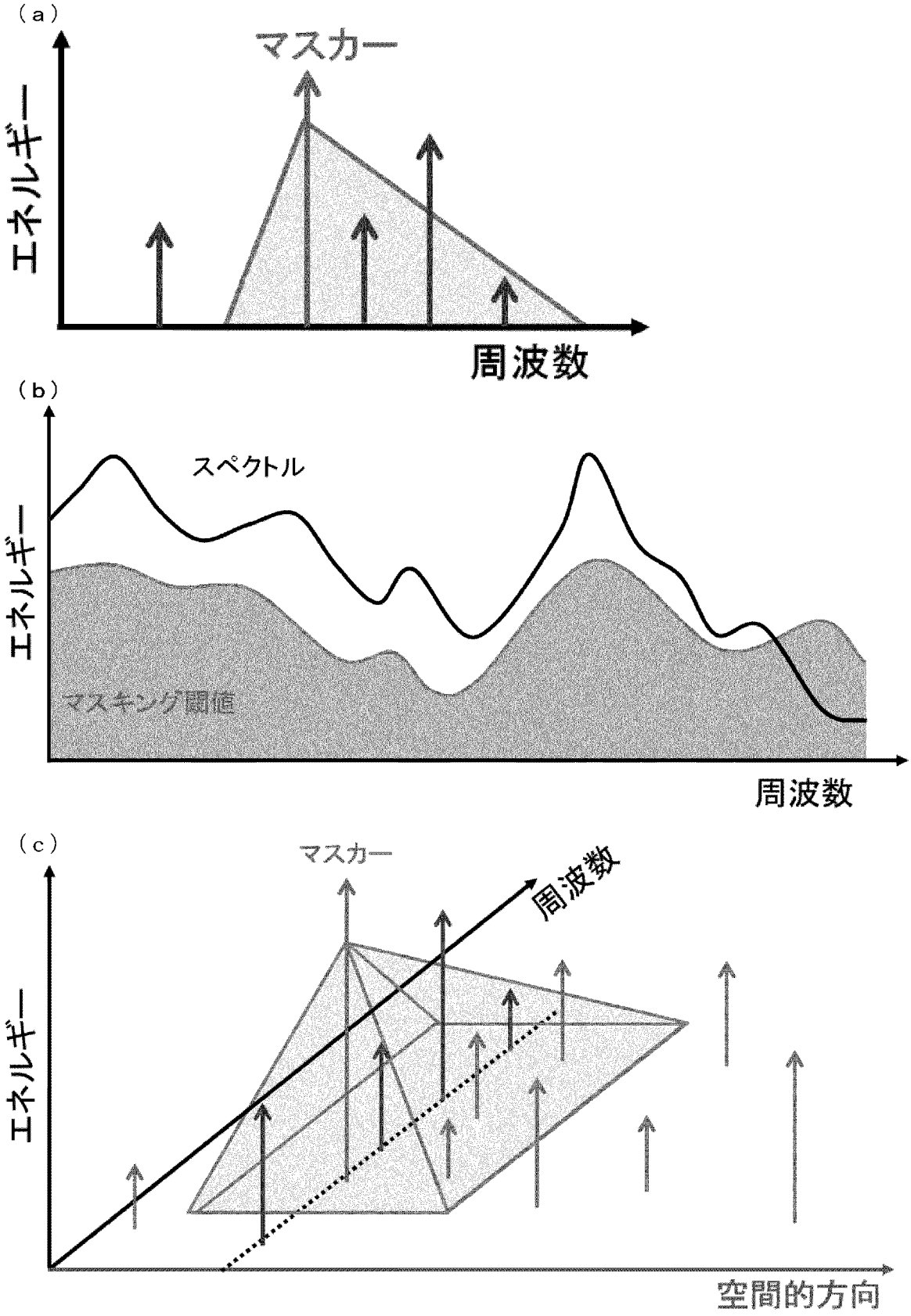
[図1]


1
符号化装置
60
符号化部
10
20
30
40
情報量
50
マイクロホンアレイ
集音部
周波数領域変換部
マスキング閾値算出部
情報量決定部
符号化伝送信号
音声信号
音響信号
マスキング閾値
音源オブジェクト
音源オブジェクトの位置情報
位置方向情報
復号化装置
2
70
80
方向算出部
送信部
90
100
110
復号化部
立体音響再生部
(L信号)
(R信号)
ヘッドフォン
音声信号
立体音響信号
X

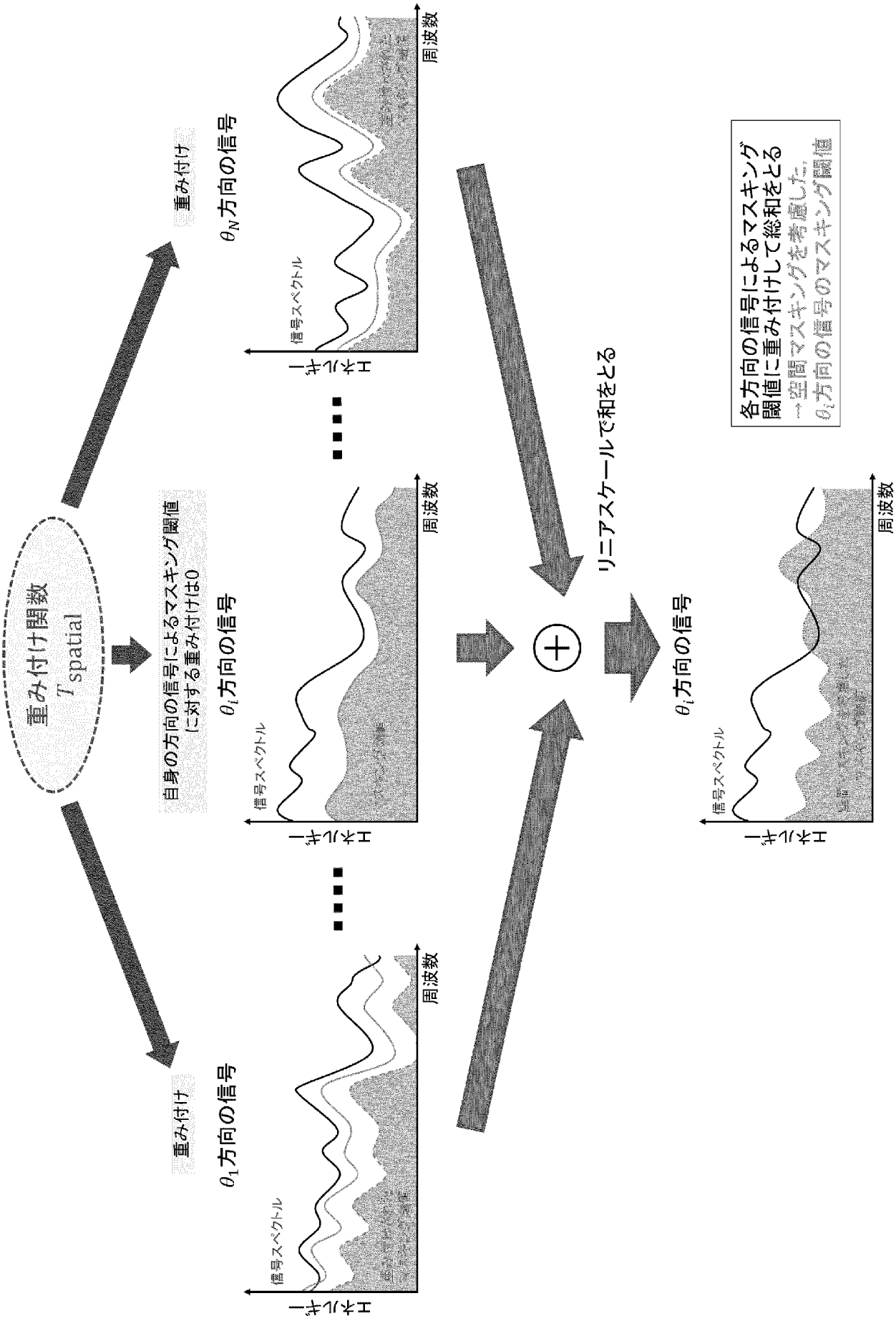
[図2]


符号化装置 1
音声データ取得処理
マスキング閾値算出処理
情報量決定処理
符号化処理
エンド
S101
S102
S103
S104
位置方向情報
符号化伝送信号
復号化装置 2
方向算出処理
方向送信処理
復号化処理
立体音響再生処理
エンド
S201
S202
S203
S204

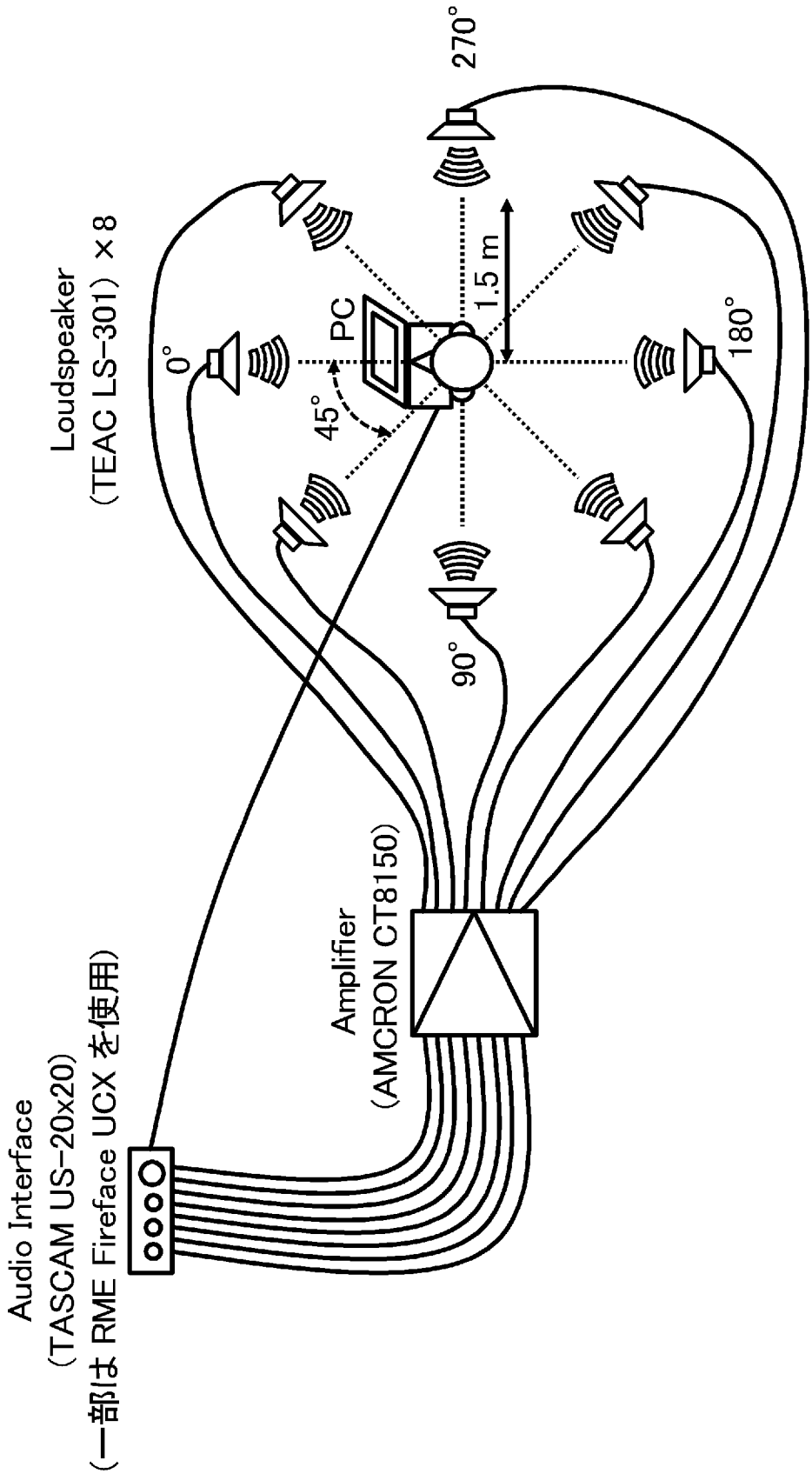
[図3]


(a)
マスカー
エネルギー
周波数
(b)
スペクトル
エネルギー
マスキング閾値
周波数
(c)
マスカー
周波数
エネルギー
空間的方向

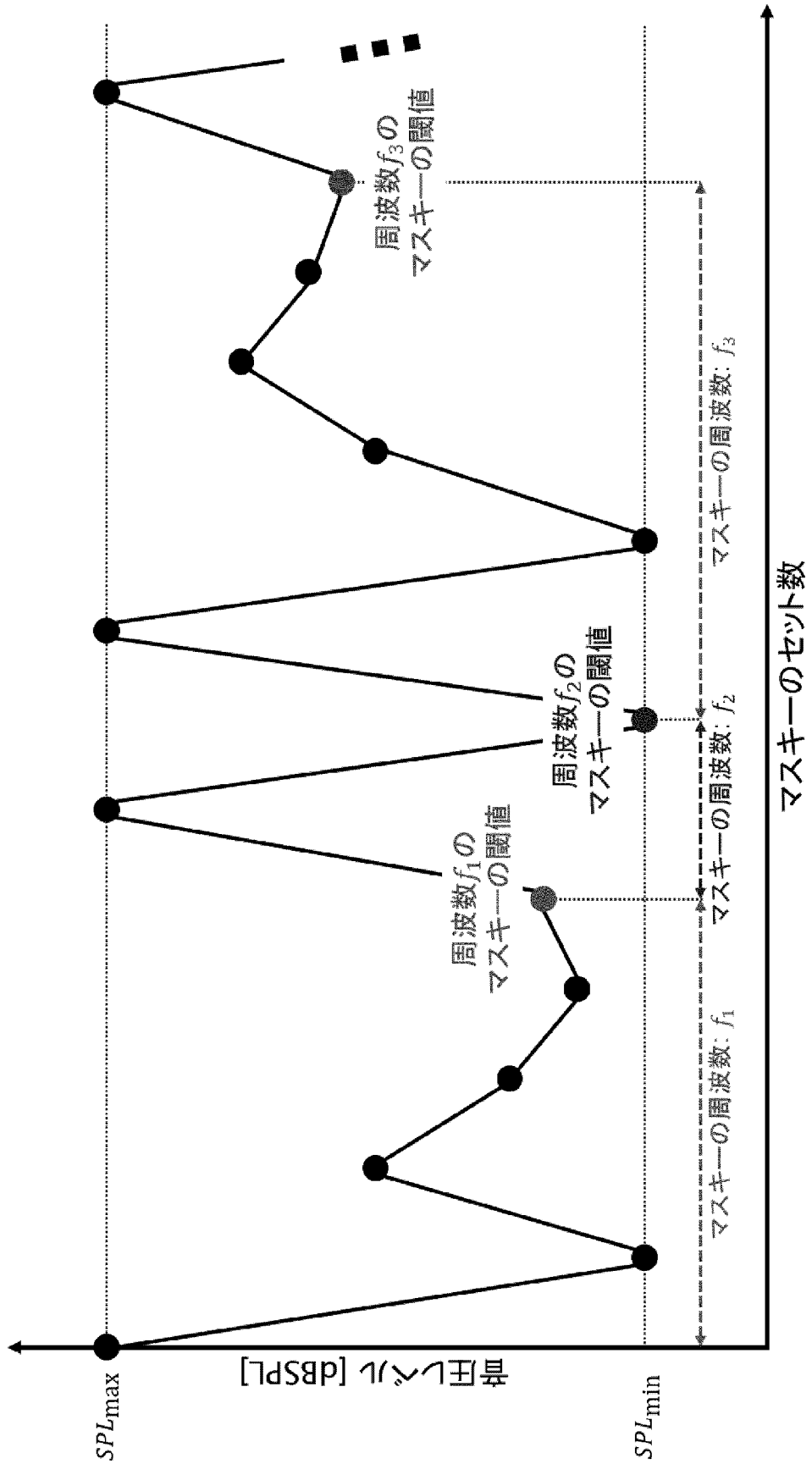
[図4]


重み付け関数
$T_{spatial}$
重み付け
重み付け
自身の方向の信号によるマスキング閾値
に対する重み付けは0
θ_1方向の信号
θ_i方向の信号
θ_N方向の信号
信号スペクトル
パワー
周波数
リニアスケールで和をとる
各方向の信号によるマスキング
閾値に重み付けして総和をとる
→空間マスキングを考慮した，
θ_i方向の信号のマスキング閾値

[図5]

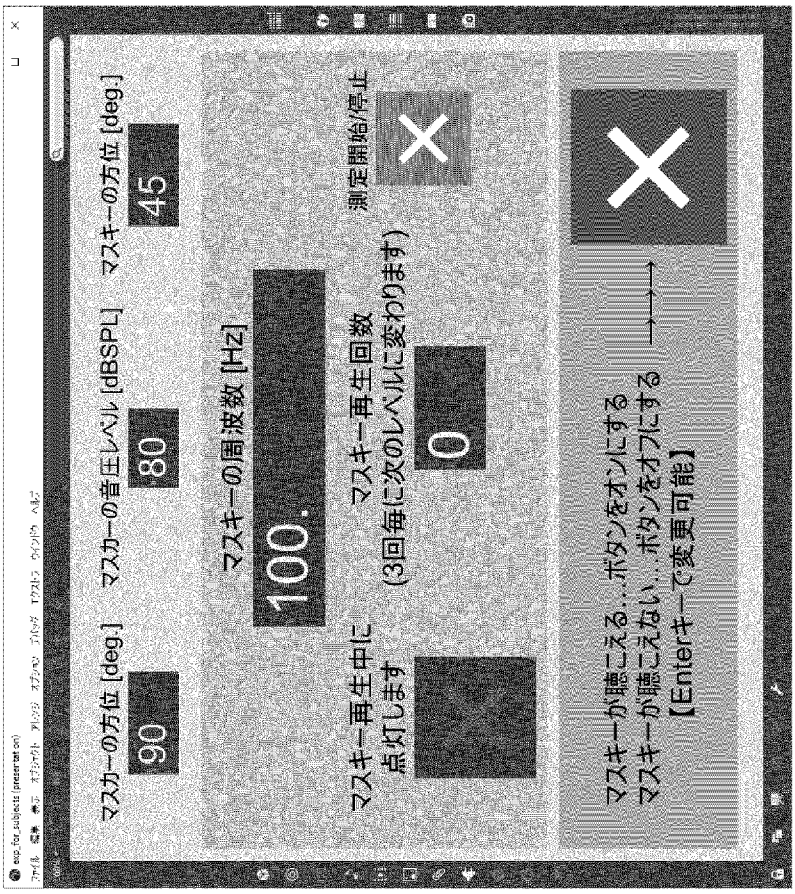

Loudspeaker
(TEAC LS-301) ×8
Audio Interface
(TASCAM US-20x20)
(一部は RME Fireface UCX を使用)
Amplifier
(AMCRON CT8150)
PC
1.5 m
0°
45°
90°
180°
270°

[図6]

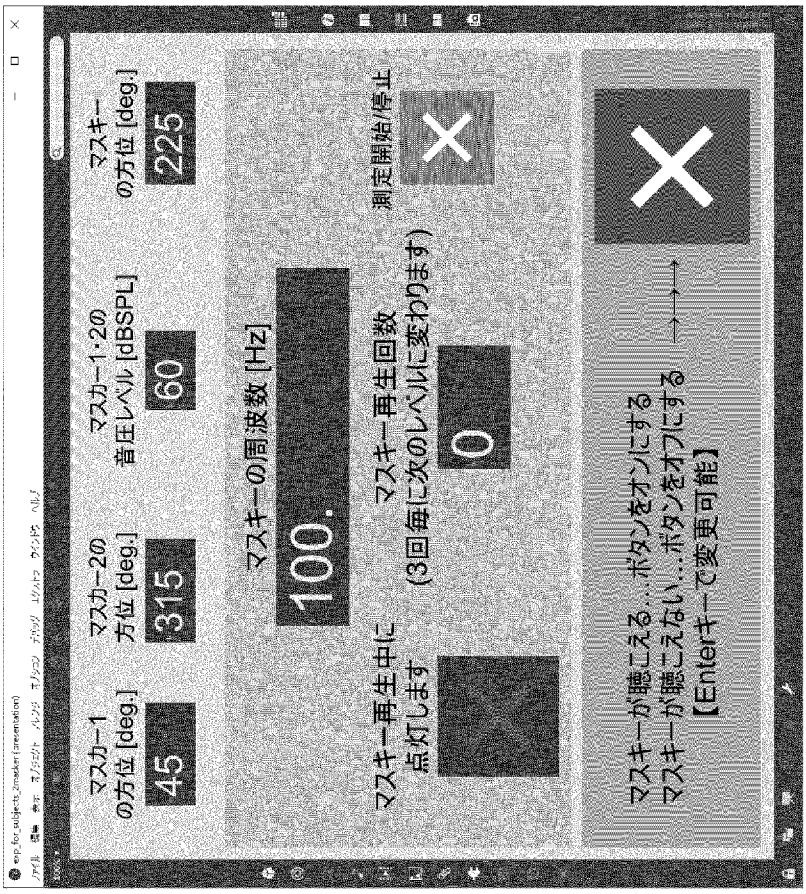

音圧レベル [dBSPL]
SPL_{max}
SPL_{min}
マスキーのセット数
周波数f_1のマスキーの閾値
周波数f_2のマスキーの閾値
周波数f_3のマスキーの閾値
マスキーの周波数: f_1
マスキーの周波数: f_2
マスキーの周波数: f_3

[図7]

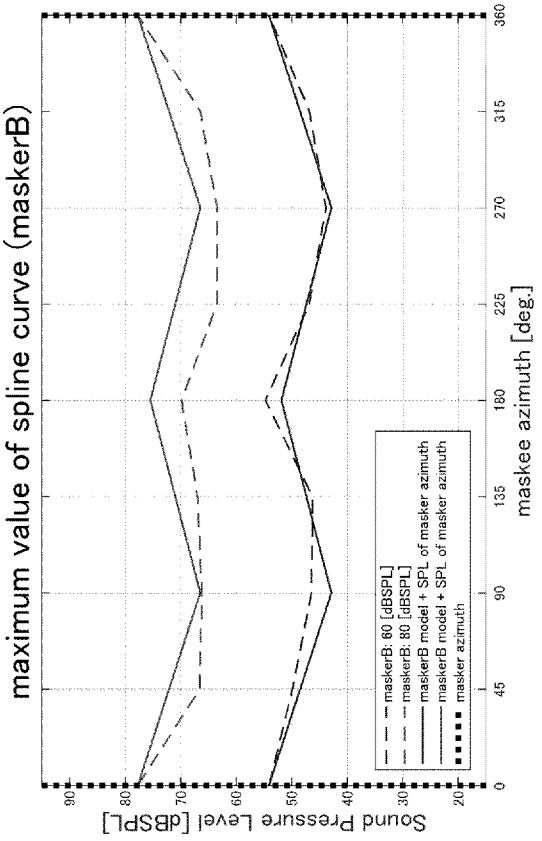
(a) マスカーが 1 音源のとき

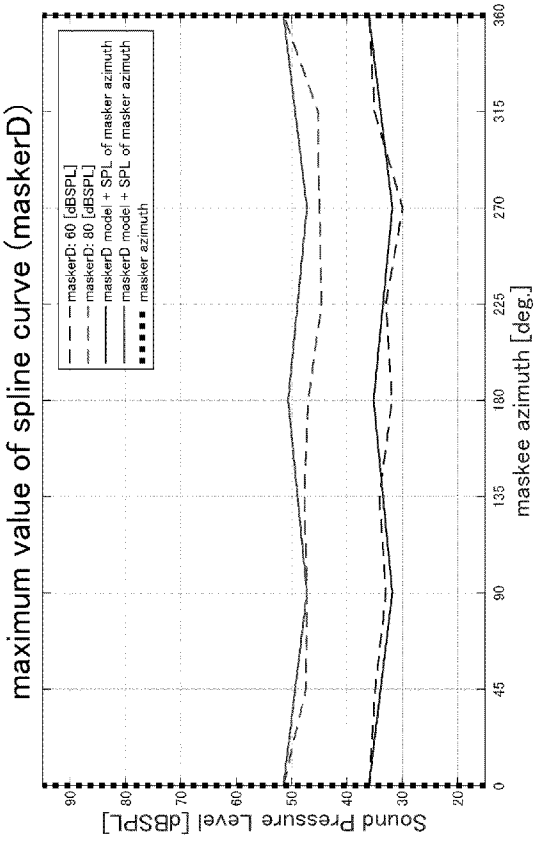


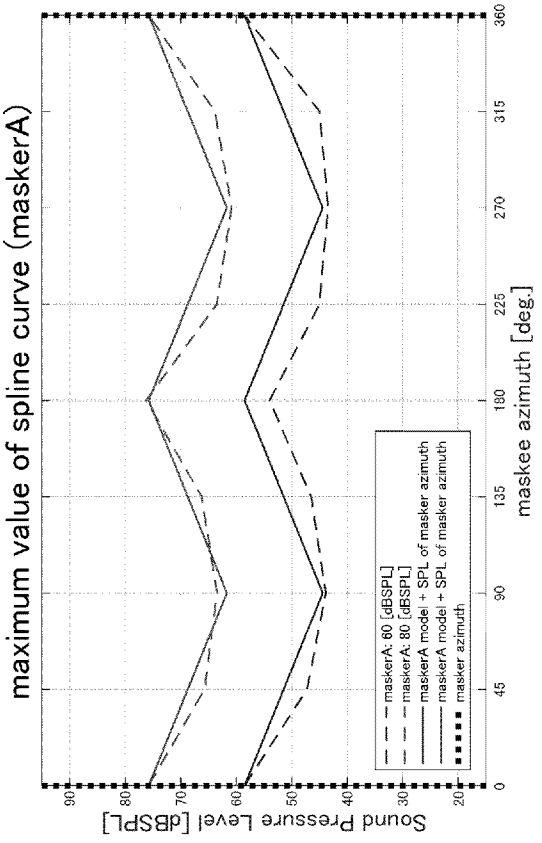
(b) マスカーが 2 音源のとき

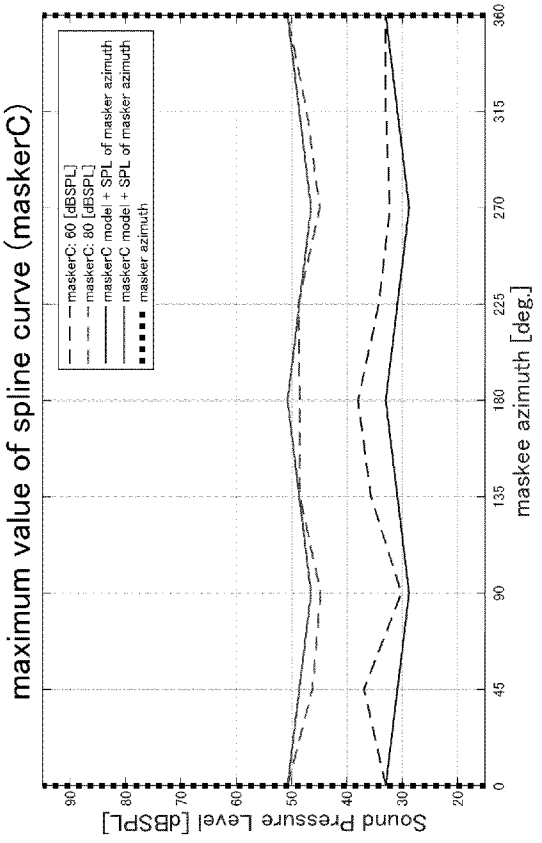


[図8]

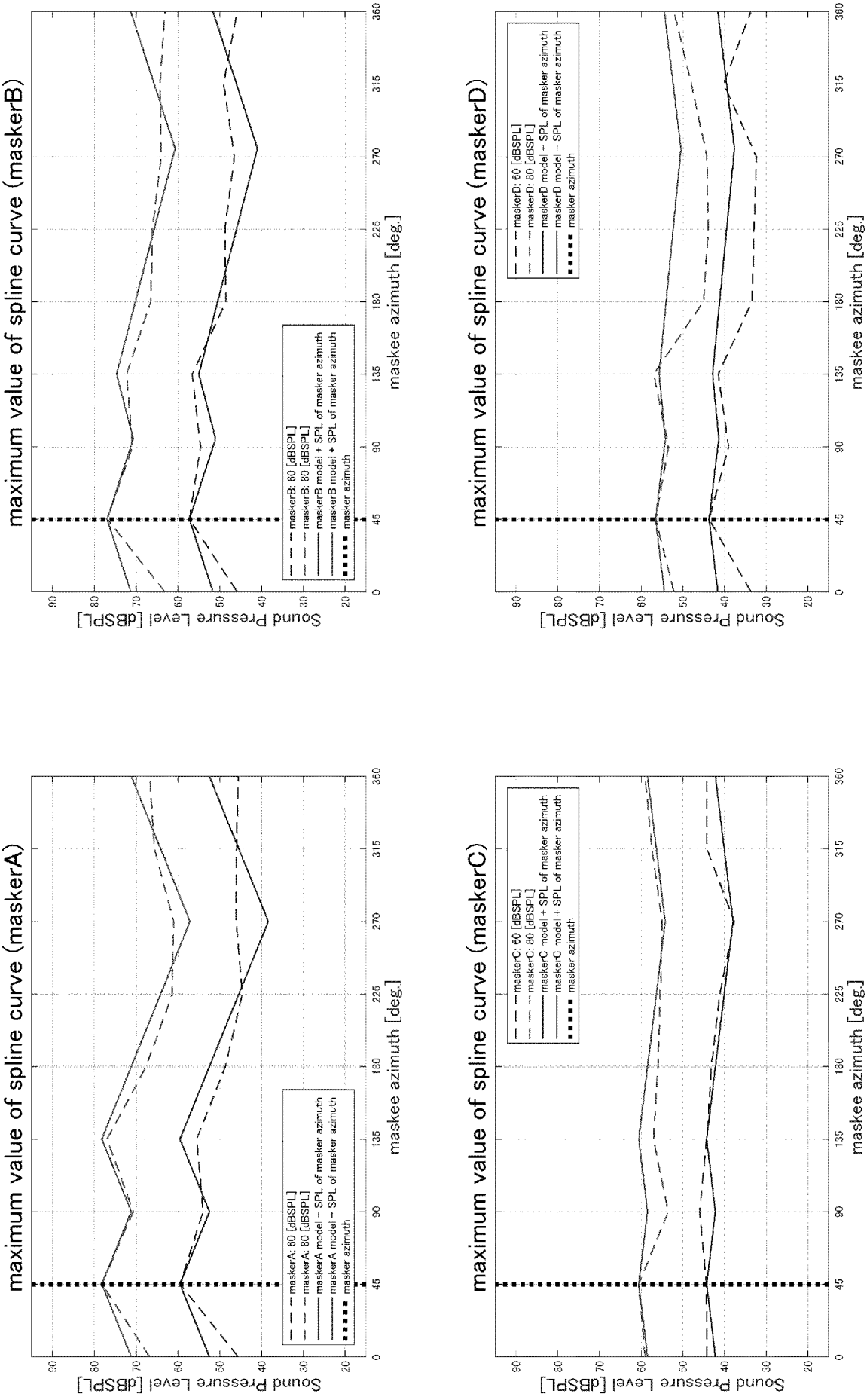

maximum value of spline curve (maskerB)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerB: 60 [dBSPL]
maskerB: 80 [dBSPL]
maskerB model + SPL of masker azimuth
maskerB model + SPL of masker azimuth
masker azimuth


maximum value of spline curve (maskerD)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerD: 60 [dBSPL]
maskerD: 80 [dBSPL]
maskerD model + SPL of masker azimuth
maskerD model + SPL of masker azimuth
masker azimuth

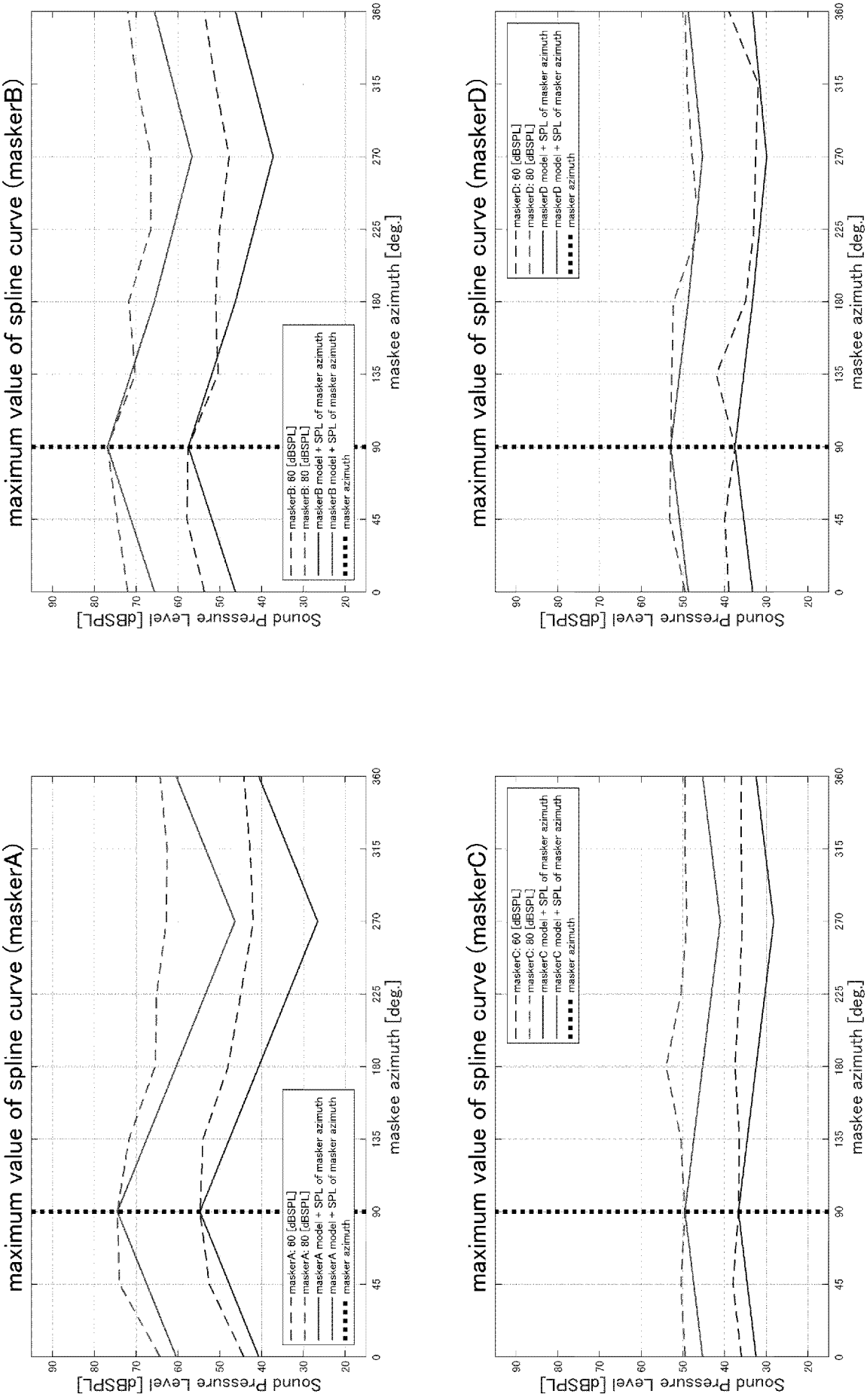

maximum value of spline curve (maskerA)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerA: 60 [dBSPL]
maskerA: 80 [dBSPL]
maskerA model + SPL of masker azimuth
maskerA model + SPL of masker azimuth
masker azimuth


maximum value of spline curve (maskerC)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerC: 60 [dBSPL]
maskerC: 80 [dBSPL]
maskerC model + SPL of masker azimuth
maskerC model + SPL of masker azimuth
masker azimuth

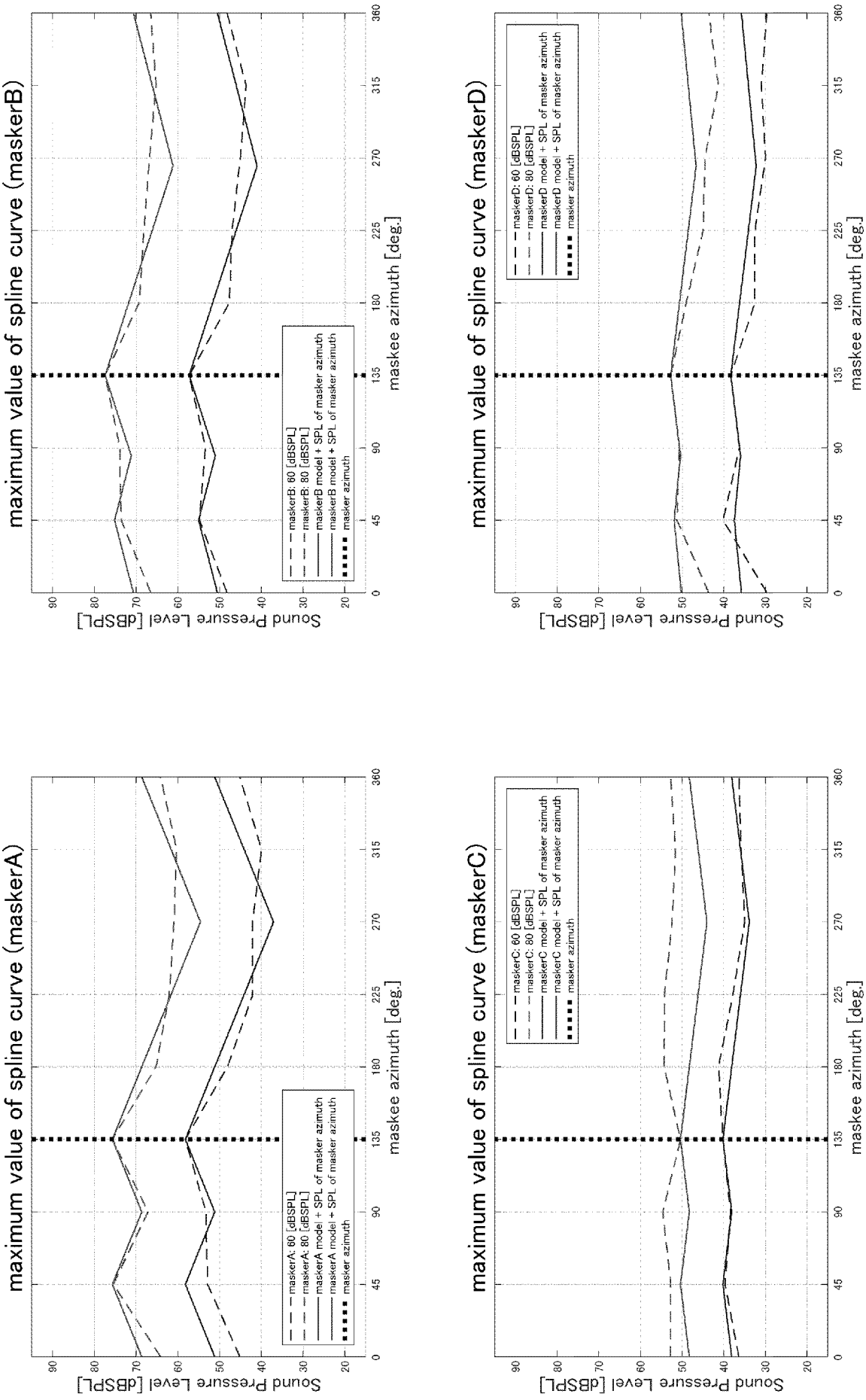
[図9]


maximum value of spline curve (maskerB)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerB: 60 [dBSPL]
maskerB: 80 [dBSPL]
maskerB model + SPL of masker azimuth
maskerB model + SPL of masker azimuth
masker azimuth
maximum value of spline curve (maskerD)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerD: 60 [dBSPL]
maskerD: 80 [dBSPL]
maskerD model + SPL of masker azimuth
maskerD model + SPL of masker azimuth
masker azimuth
maximum value of spline curve (maskerA)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerA: 60 [dBSPL]
maskerA: 80 [dBSPL]
maskerA model + SPL of masker azimuth
maskerA model + SPL of masker azimuth
masker azimuth
maximum value of spline curve (maskerC)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerC: 60 [dBSPL]
maskerC: 80 [dBSPL]
maskerC model + SPL of masker azimuth
maskerC model + SPL of masker azimuth
masker azimuth

[図10]


maximum value of spline curve (maskerB)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerB: 60 [dBSPL]
maskerB: 80 [dBSPL]
maskerB model + SPL of masker azimuth
maskerB model + SPL of masker azimuth
masker azimuth
maximum value of spline curve (maskerD)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerD: 60 [dBSPL]
maskerD: 80 [dBSPL]
maskerD model + SPL of masker azimuth
maskerD model + SPL of masker azimuth
masker azimuth
maximum value of spline curve (maskerA)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerA: 60 [dBSPL]
maskerA: 80 [dBSPL]
maskerA model + SPL of masker azimuth
maskerA model + SPL of masker azimuth
masker azimuth
maximum value of spline curve (maskerC)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerC: 60 [dBSPL]
maskerC: 80 [dBSPL]
maskerC model + SPL of masker azimuth
maskerC model + SPL of masker azimuth
masker azimuth

[図11]


maximum value of spline curve (maskerB)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerB: 60 [dBSPL]
maskerB: 80 [dBSPL]
maskerB model + SPL of masker azimuth
maskerB model + SPL of masker azimuth
masker azimuth
maximum value of spline curve (maskerD)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerD: 60 [dBSPL]
maskerD: 80 [dBSPL]
maskerD model + SPL of masker azimuth
maskerD model + SPL of masker azimuth
masker azimuth
maximum value of spline curve (maskerA)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerA: 60 [dBSPL]
maskerA: 80 [dBSPL]
maskerA model + SPL of masker azimuth
maskerA model + SPL of masker azimuth
masker azimuth
maximum value of spline curve (maskerC)
Sound Pressure Level [dBSPL]
maskee azimuth [deg.]
maskerC: 60 [dBSPL]
maskerC: 80 [dBSPL]
maskerC model + SPL of masker azimuth
maskerC model + SPL of masker azimuth
masker azimuth

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2020/006211

A. CLASSIFICATION OF SUBJECT MATTER
G10L 19/002(2013.01)i; G10L 19/008(2013.01)i; G10L 19/035(2013.01)i
FI: G10L19/035 E; G10L19/002; G10L19/008 200

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L19/00-19/26

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2020
Registered utility model specifications of Japan	1996-2020
Published registered utility model applications of Japan	1994-2020

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
IEEE Xplore

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	JP 2016-524726 A (QUALCOMM INCORPORATED) 18.08.2016 (2016-08-18) paragraphs [0061]-[0068], [0076]-[0080]	1,3-4,7,9-13,15-16,18 5-6,8,14,17
X A	JP 2015-531078 A (INTELLECTUAL DISCOVERY CO., LTD.) 29.10.2015 (2015-10-29) paragraphs [0049], [0054], [0070]-[0077]	2-4,7,9,12-13,15-16,18 5-6,8,14,17
A	JP 2016-518788 A (DOLBY LABORATORIES LICENSING CORPORATION) 23.06.2016 (2016-06-23) paragraph [0031]	7
A	JP 2014-016625 A (DOLBY INTERNATIONAL AB) 30.01.2014 (2014-01-30) paragraph [0143]	9

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search 12 June 2020 (12.06.2020)	Date of mailing of the international search report 30 June 2020 (30.06.2020)
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2020/006211

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2016-224472 A (DOLBY INTERNATIONAL AB) 28.12.2016 (2016-12-28) entire text, all drawings	1-18
P,X P,A	加藤巧大ほか，聴覚の空間的マスキング効果を考慮した3D音響信号の符号化に関する基礎検討，電子情報通信学会技術研究報告，07 March 2019, vol. 118, no. 497, pp. 271-278, ISSN 2432-6380, in particular, chapter 3-4, (KATO, Kodai et al., "Study on 3D audio coding based on spatial auditory masking", IEICE technical report)	1-9,11-12 10,13-18

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2020/006211

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2016-524726 A	18 Aug. 2016	US 2014/0355768 A1 paragraphs [0059]-[0066], [0074]-[0078] CN 105247612 A KR 10-2016-0012215 A	
JP 2015-531078 A	29 Oct. 2015	US 2015/0194158 A1 paragraphs [0070], [0076], [0094]-[0100] KR 10-2014-0017342 A CN 104541524 A	
JP 2016-518788 A	23 Jun. 2016	US 2016/0072467 A1 paragraph [0038] CN 105164918 A	
JP 2014-016625 A	30 Jan. 2014	EP 2077550 A1 paragraph [0134] KR 10-2010-0105745 A CN 101925950 A	
JP 2016-224472 A	28 Dec. 2016	US 2012/0155653 A1 entire text, all drawings CN 102547549 A KR 10-2012-0070521 A	

国際調査報告

国際出願番号
PCT/JP2020/006211

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G10L 19/002(2013.01)i; G10L 19/008(2013.01)i; G10L 19/035(2013.01)i
FI: G10L19/035 E; G10L19/002; G10L19/008 200

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

G10L19/00-19/26

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０２０年
日本国実用新案登録公報	１９９６－２０２０年
日本国登録実用新案公報	１９９４－２０２０年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

IEEE Xplore

C. 関連すると認められる文献

引用文献のカテゴリー*	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
X	JP 2016-524726 A (クゥアルコム・インコーポレイテッド) 18.08.2016 (2016-08-18) 段落[0061]-[0068],[0076]-[0080]	1,3-4,7,9-13,15-16,18
A		5-6,8,14,17
X	JP 2015-531078 A (インテレクチュアル ディスカバリー シーオー エルティディ) 29.10.2015 (2015-10-29) 段落[0049],[0054],[0070]-[0077]	2-4,7,9,12-13,15-16,18
A		5-6,8,14,17
A	JP 2016-518788 A (ドルビー ラボラトリーズ ライセンシング コーポレイション) 23.06.2016 (2016-06-23) 段落[0031]	7
A	JP 2014-016625 A (ドルビー・インターナショナル・アーベー) 30.01.2014 (2014-01-30) 段落[0143]	9

☑ Ｃ欄の続きにも文献が列挙されている。

☑ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー
"A" 特に関連のある文献ではなく、一般的技術水準を示すもの
"E" 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
"L" 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
"O" 口頭による開示、使用、展示等に言及する文献
"P" 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
"T" 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの
"X" 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
"Y" 特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
"&" 同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
12.06.2020	30.06.2020

名称及びあて先

日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）

中村　天真 5Z 1786

電話番号 03-3581-1101 内線 3591

C.　関連すると認められる文献

引用文献のカテゴリー*	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
A	JP 2016-224472 A (ドルビー・インターナショナル・アーベー) 28.12.2016 (2016-12-28) 全文，全図	1-18
P，X	加藤巧大ほか，聴覚の空間的マスキング効果を考慮した３Ｄ音響信号の符号化に関する基礎検討，電子情報通信学会技術研究報告，2019.03.07，第118巻，第497号，p.271-278，ISSN 2432-6380 特に3-4章	1-9,11-12
P，A		10,13-18

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2020/006211

引用文献	公表日	パテントファミリー文献	公表日
JP 2016-524726 A	18.08.2016	US 2014/0355768 A1 段落[0059]-[0066],[0074]-[0078] CN 105247612 A KR 10-2016-0012215 A	
JP 2015-531078 A	29.10.2015	US 2015/0194158 A1 段落[0070],[0076],[0094]-[0100] KR 10-2014-0017342 A CN 104541524 A	
JP 2016-518788 A	23.06.2016	US 2016/0072467 A1 段落[0038] CN 105164918 A	
JP 2014-016625 A	30.01.2014	EP 2077550 A1 段落[0134] KR 10-2010-0105745 A CN 101925950 A	
JP 2016-224472 A	28.12.2016	US 2012/0155653 A1 全文，全図 CN 102547549 A KR 10-2012-0070521 A	