

# 任意の印象に合致した音声表現を持つ言葉の自動構成方式

本間 秀典<sup>†</sup> 中西 崇文<sup>††</sup> 北川 高嗣<sup>†††</sup>

<sup>†</sup> 筑波大学大学院システム情報工学研究科 〒305-8573 茨城県つくば市天王台 1-1-1

<sup>††</sup> 情報通信研究機構 〒619-0289 京都府相楽郡精華町光台 3-5

<sup>†††</sup> 筑波大学大学院システム情報工学研究科 〒305-8573 茨城県つくば市天王台 1-1-1

E-mail: †homma@mma.cs.tsukuba.ac.jp, ††takafumi@nict.go.jp, †††Takashi@cs.tsukuba.ac.jp

**あらまし** 本稿では、任意の言葉を印象語として入力することにより、その印象に合致した音声表現を自動構成するための方式を提案する。本方式では、音相理論と呼ばれる理論を用いることにより、言葉の音声表現と印象語の相互の変換を実現する。さらに、本方式を意味の数学モデルに適用することにより、言葉の音声表現と任意の印象語、およびそれにより表現される任意のメディアデータを自由に連結することが可能となると考えられる。

**キーワード** 意味の数学モデル, メタデータ, 音相, 感性, 逆演算

## A Method of Automatic Composition of Sound Expression of A Word Corresponding to the Impressions of Arbitrary Words

Hidenori HOMMA<sup>†</sup>, Takafumi NAKANISHI<sup>††</sup>, and Takashi KITAGAWA<sup>†††</sup>

<sup>†</sup> Graduate School of Systems and Information Engineering, University of Tsukuba  
1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573 Japan

<sup>††</sup> National Institute of Information and Communications Technology  
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289 Japan

<sup>†††</sup> Graduate School of Systems and Information Engineering, University of Tsukuba  
1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573 Japan

E-mail: †homma@mma.cs.tsukuba.ac.jp, ††takafumi@nict.go.jp, †††Takashi@cs.tsukuba.ac.jp

**Abstract** In this paper, we present an automatic composition method of sound expression of a word corresponding to the impressions of arbitrary words. We have proposed a method of automatic metadata extraction corresponding to the impression of the sounds of arbitrary words by the research called “Onso”, and we propose an inverse operation of the automatic metadata extraction in this paper. This method enables a connection of any words and mediadata represented in impression words.

**Key words** Mathematical model of meaning, Metadata, Onso, Kansei, Inverse operation

### 1. まえがき

近年、コンピュータネットワークの広域化・高速化やコンピュータの廉価化により、多種多様なメディア情報群がネットワーク上に散在しつつある。コンピュータを介してこれらの情報群と接触する機会は増大する傾向にあり、これらのメディア情報を対象とした、より直感的な情報検索方式、及び知識の獲得方式の確立が重要となってきた。

一般に、我々のコミュニケーションにおいて、互いの感情を正確に理解することが重要である。メディアデータ群を対象としたシステム環境において、感性的な情報の伝達が可能な機構が人間とシステムとの間のコミュニケーションメディアとし

て導入されれば、人間の感性や直感に合致した、ユーザへの負荷がより少ないインターフェースの実現が可能となると考えられる。

文献[1],[2]によれば、一般に語や語句はさまざまな印象を喚起するが、語の印象はその語の持つ意味からだけでなく、語の音の特徴からも喚起されるとしている。日本語における言葉の音とその印象の相関関係は、「音相理論」[1],[2]と呼ばれる研究の中で示されている。音相は、我々の日常的なコミュニケーションにおいて、互いの感情を理解するための重要な要素のひとつである。

これまで我々は、文献[3],[4]でこの音相理論に基づいて任意の言葉の音声表現を持つ印象語をメタデータとして自動抽出す

る方式を提案している。これにより、任意の言葉をコンテキストに用いた、感性情報をより直観的にコンピュータに伝達可能な連想検索インターフェースの実現が可能となった。

本稿では、このメタデータの自動抽出方式の逆演算を定義することにより、任意の印象語を対象として、音相理論によりその印象に合致した音声表現を持つ言葉を自動的に構成する方式を提案する。提案方式がコミュニケーションメディアとして実装されることにより、ユーザとシステムとの間の相互の感性情報の伝達が可能となると考えられる。さらに、複数の異種のメディア情報の連結が可能となることから、人間のマルチモーダル性を生かした情報の検索や呈示などの新たな機能の実現が可能となると考えられる。

音相に代表される、感情などの様々な情報を効果的に表すメディアを含む異種のメディアを、意味の数学モデルに代表される連想検索方式に適用することによって統一的かつ双方向に扱うことが可能な方式を実現できれば、人間の感性や直感に合致した、マルチメディア環境での有効なコミュニケーションメディアの実現が可能となる。さらに、これにより独立に存在するメディアデータ群やシステム群の相互運用性を高め、新たな価値の創造が可能になると考えられる。

## 2. 意味の数学モデルの概要

本稿では、言葉と言葉の意味的な関係の計量を実現する、意味の数学モデルの概要を示す。詳細は文献 [5]~[7] に述べられている。

### (1) メタデータ空間 *MDS* の設定

検索対象となるメディアデータをベクトルで表現したデータにマッピングするための正規直交空間（以下、メタデータ空間 *MDS*）を設定する。

### (2) メディアデータのメタデータをメタデータ空間 *MDS* へ写像

メディアデータのメタデータをベクトル化し、メタデータ空間 *MDS* へ写像する。これにより、各検索対象データのメタデータが同じメタデータ空間上に配置されることになり、検索対象データ間の意味的な関係を空間上での距離として計算することが可能となる。

(3) メタデータ空間 *MDS* の部分空間（意味空間）の選択  
検索者は与える文脈を複数の単語を用いて表現する。検索者が与える単語の集合をコンテキストと呼ぶ。各コンテキストに対応するベクトルをメタデータ空間 *MDS* に写像し、これらのベクトルをメタデータ空間 *MDS* 上で合成することにより、意味重心を表すベクトルが生成される。意味重心から各軸への射影値を相関値とし、閾値を超えた相関値（以下、重み）を持つ軸からなる部分空間（以下、意味空間）が選択される。

### (4) メタデータ空間 *MDS* の部分空間（意味空間）における相関の定量化

選択されたメタデータ空間 *MDS* の部分空間（意味空間）において、メディアデータベクトルのノルムをそのメディアデータの検索語列との相関として計量する。これにより、与えられたコンテキストと各メディアデータとの相関の強さを定量化して

いる。この意味空間における検索結果は、各メディアデータを相関の強さについてソートしたリストとして与えられる。

## 3. 任意の印象に合致した音声表現を持つ言葉の自動構成方式

本節では、入力された任意の言葉を対象とした、その印象と合致する音を持つ言葉の自動生成方式を提案する。提案方式は、音相理論に基づいて入力された言葉の音を持つ印象に合致した印象語をメタデータとして抽出する方式の逆演算として定義される。

3.1 節では、提案方式の基礎概念である音相理論の概要を述べる。3.2 節では、任意の言葉を対象とした音の印象によるメタデータ自動抽出方式について概説する。3.3 節では、メタデータ自動抽出方式の逆演算を定義することにより、任意の言葉の印象に合致した言葉の音声表現の自動構成方式を提案する。

### 3.1 音相理論

文献 [1],[2] によれば、「音相」とは、言葉の音の響きを持つ印象であり、同じ言葉を話すほとんどの話者がその言葉に対して共通して持っているイメージであるといえる。音相理論によれば、我々は、その言葉のもつ意味の微妙な違いを、似た意味の語の中から、音相に基づいてよりそのイメージに近い音を持つ語を選んで使うことによっても伝えているとされる。そのため音相は、我々が言葉によって伝えようとしている感情やイメージなどを直感的に表現していると考えられる。

音相理論では、音相基と呼ばれる音相を構成する要素を用いて対象となる語の印象を決定している。音相基の中でも特に、それぞれの母音や子音が発音される時に用いられる器官（唇、舌、喉など）とそれらの操作の仕方にあたる「調音種」と、それぞれの音がどのような感情のときに用いられるかという2つの観点から音相による語の表情が作られるとされている。後者に関しては、その音の明るさを表す「輝性」(B, Brightness の略) および強さを表す「動性」(H, Hardness の略) という要素を単位としている。それぞれの子音や母音の調音種とそれらが持つ輝性、動性の関係は文献 [2] において表 1, 2 のように示されている。各表において、B の値は符号付きで表され、正であれば語の音の印象が明るく、負であれば印象が暗いことを示し、絶対値の大きさはその度合いを示している。また、H の値は符号無しの数値で表され、数値が大きければ語の音が強いインパクトを持っていることを表している。

文献 [2] では、音相による表情を表 3 に示すように 20 の群と 40 の表情属性（語）に分類している。さらに、音相を音相基それぞれが持つ表情である甲類表情と、幾つかの音相基の組み合わせによって生まれる乙類表情の 2 種類に分類している。甲乙それぞれの表情と表情語の関係は表 4, 5 のように示されている。

### 3.2 任意の言葉を対象とした音の印象によるメタデータ自動抽出方式

本節では、任意の言葉を対象とした音の印象によるメタデータ自動抽出方式について述べる。本方式は、音相理論に基づいて任意の言葉の音（発音）から受ける印象をメタデータとして自動抽出する方式である。本方式は以下の手順により実現さ

表1 音素と音価表(子音の場合).

Table 1 The table of Onso for consonants.

		両唇音	前舌音	咽頭音
破裂音	無	p	t	k
	有	+B1.7 H1.6	+B1.4 H1.5	+B1.3 H1.3
摩擦音	無	b	d	g
	有	-B2.2 H1.2	-B2.0 H1.0	-B2.0 H1.0
破裂音	無		ts	tʃ
	有		+B1.7 H2.2	+2.2 H2.7
摩擦音	無	f	s	ʃ
	有	B0.0 H0.0	+B0.3 H0.7	+B0.8 H1.2
弾き音 (流音)	無		r	
	有		+B0.4 H0.8	
鼻音	無	m	n	ŋ
	有	B0.0 H0.0	B0.0 H0.0	B0.0 H0.0
接近音	無	w	j	
	有	-B0.5 H0.0	+B0.5 H0.5	

- ・ 無：無声音
- ・ ts：ツの子音
- ・ tʃ：チおよびチャ行の子音
- ・ f：フおよびファ行の子音
- ・ ʃ：シおよびシャ行の子音
- ・ ɲ：ガ行鼻音
- ・ j：ヤ行子音および拗音の第2子音
- ・ 促音(Q)は後続子音に H1.3 を加える
- ・ 拗音は前子音に +B0.5, H0.5 をそれぞれ加える
- ・ 撥音(N)は -B0.4 H0.0

表2 音素と音価表(母音の場合).

Table 2 The table of Onso for vowels.

a	有声音	B0.0	H0.0
i		+B1.0	H1.0
u		-B1.0	H0.0
e		B0.0	H0.0
o		-B0.7	H0.0

れる。

(1) 入力語からの初期ベクトル  $v_b$  の生成

入力語から甲類表情を抽出するための基本要素として、表6に示す29の要素を抽出し、これらを要素とする29次元の初期ベクトル  $v_b$  を生成する。

(2) 甲類表情に対応するベクトル  $v'_k$  の生成

行要素に表6に示された29種の音相基を、列要素に表4に示された40種の甲類表情をそれぞれ配置して特徴づけを行うことにより、任意の言葉から抽出された音相基ベクトル  $v_b$  から甲類表情に対応したベクトル  $v'_k$  へ変換するための変換行列  $T_1$  を作成する。ここでは、ある甲類表情を生じる音相基に対応する要素の値を“1”、それ以外の要素を“0”とすることによって  $T_1$  の特徴づけを行う。そして、以下の演算により、 $v_b$  を40種の甲類表情に対応するベクトル  $v'_k$  に変換する。

$$v'_k = T_1 v_b$$

表3 表情語と表情属性表(抜粋).

Table 3 The table of impression words.

表情語群	表情属性	訳語
A	シンプルな, 明白さ	plain, obvious
B	躍動感, 進歩的	vibrant, advance
C	新鮮さ, 新奇さ	fresh, unprecedented
D	動的, 活性的	dynamic, active
E	派手さ, 賑やかさ	florid, bustle
F	軽やかさ, 軽快感	light, trippingly
G	若さ, 澆刺さ	young, effervescent
H	現代的, 都会的	modern, urban
I	明るさ, 開放的	bright, open-minded
J	合理的, 現実的	reasonable, real
K	個性的, 特殊的	individual, special
L	強さ, 鋭さ	powerful, sharp
M	適応性, 庶民的	adaptable, popular
N	清らかさ, 爽やかさ	pure, brisk
O	健康的, 清潔感	healthy, clean
P	暖かさ, 安らぎ	warm, comfortable
Q	安定感, 信頼感	stable, confidence
R	高級感, 充実感	expensive, fulfil
S	高尚な, 優雅さ	profound, elegant
T	静的, 非活性的	static, inactive

$$= (x_1, x_2, \dots, x_{40})^T.$$

(3) 標準使用率の適用による  $v'_k$  の甲類表情を表すベクトル  $v_k$  への変換

音相理論では、ある音相基（もしくは、幾つかの音相基の和）の値と、表7に示された言葉全体の拍数に対する各音相基の標準使用率から、それに対応する甲類表情がどの程度表れてくるかを判定している。文献[2]によれば、その音相基を「多用している」及び「過少である」ことの基準として、標準使用率  $\pm 15$  ポイントが挙げられている。そこで、本稿では、 $N$  拍の音を持つある言葉から抽出される  $v_k$  の  $i$  番目の要素を  $x_i$ 、及び音相基の標準値を  $n_i$  とした場合の、その甲類表情との相関の値  $p_i$  を式(1)により甲類表情の判定演算を行うこととした。

$$p_i = \frac{1}{0.15} \times \left( \frac{x_i}{N} - n_i \right). \quad (1)$$

この  $p_i$  を用いることによって、 $v'_k$  を甲類表情を表すベクトル  $v_k$  に変換する。

$$v_k = (p_1, p_2, \dots, p_{40})^T.$$

(4)  $v_k$  の甲類・乙類表情を表すベクトル  $v_f$  への拡張

$v_k$  の各要素と表5を用いることによって、 $v_k$  を乙類表情を反映した78次元のベクトル  $v_f$  に変換する。

$$v_f = (p_1, p_2, \dots, p_{40}, a_1, a_2, \dots, a_{38}).$$

ここで、 $a_k (k = 1, 2, \dots, 38)$  は、表5の  $k$  番目に示された乙類表情に対応する甲類表情の組  $(x_i, x_j)$  を用いて以下のように表される。

表4 甲類表情.

Table 4 Individual properties.

音相基	抽出される表情語群
1. ア音多用	I,M,T,P
2. イ音多用	A,G,I,K,L,O,P
3. ウ, オ音多用	Q,S,T
4. エ音多用	I,K,M,P
5. 母音種多用	B,M,P,G,E,I,O
6. 母音種過少	C,K
7. 有声音多用	P,Q,S,T
8. 無声音多用	D,I,H,O,R
9. 有声破裂音系多用	Q,R,S,T
10. 無声破裂音系多用	D,F,G,I,J,N
11. 有声摩擦音多用	Q,T
12. 無声摩擦音系多用	I,N,O
13. 鼻音多用	F,M,P
14. 流音多用	D,F
15. 高調音種比	A,B,C,E,F,G,O
16. 低調音種比	K,Q
17. R(長音)多用	Q,R,T
18. Q(促音)多用	A,B,C,D,F,I,L,O
19. N(撥音)多用	Q
20. 濁音多用	P,Q,R,S,T
21. 高動性	A,B,C,D,E,G,H,I,L
22. 低動性	P,Q,T
23. プラス輝性	C,D,F,G,H,I,N,O
24. プラス高輝性	G,H,I,K,M
25. マイナス輝性	Q,R,S
26. マイナス高輝性	Q,R,T
27. 高動輝拍多用	A,D,J,K,S
28. 順接拍多用	A,C,D,E,F,I,M,O
29. 逆接拍多用	P,Q,R,S,T
30. 多拍	P,R,S,T
31. 少拍	A,F,J,L
32. 無声化母音多用	B,C,F,G,H,I,N,O
33. 無声拗音多用	A,B,C,D,E,F,G,H,I,J
34. 子音拍多用	A,B,C,D,F,G,H,I,J,K,O,Q
35. 子音拍過少	R,S
36. 新子音多用	C,G,H
37. 高総合音価	B,C,D,E,G,H,I,M,P
38. 低総合音価	P,Q
39. 濁拗音多用	E,S,T
40. 摩擦音系多用	N,O,P,T

表5 乙類表情.

Table 5 Anaphoric properties.

甲類表情の組	抽出される表情語群
1 甲1×甲17	P, T
2 甲2×甲10	G, I
3 甲2×甲27	A, J, K
4 甲3×甲9	Q, S, T
5 甲5×甲8	I, O, P
6 甲5×甲28	E, I, M, O
7 甲6×甲16	K
8 甲6×甲27	K
9 甲6×甲29	C, K
10 甲6×甲37	C
11 甲7×甲16	Q
12 甲7×甲25	Q, S
13 甲7×甲27	S
14 甲7×甲29	P, Q, S, T
15 甲7×甲30	P, S, T
16 甲8×甲12	I, N, O
17 甲8×甲37	D, H, I, P
18 甲10×甲12	I, N
19 甲10×甲18	F, J
20 甲10×甲33	D, F, G, I, J
21 甲40×甲13	P
22 甲12×甲32	I, N, O
23 甲13×甲14	F
24 甲15×甲27	A
25 甲15×甲28	A, B, C, E, F, O
26 甲15×甲32	B, C, F, O
27 甲15×甲37	B, C, E, G
28 甲16×甲27	K
29 甲16×甲29	Q
30 甲18×甲28	A, C, F, O
31 甲25×甲29	Q, R, S
32 甲26×甲29	Q, R, T
33 甲27×甲29	S
34 甲28×甲32	B, C, F, I, O
35 甲29×甲37	P
36 甲32×甲34	B, C, F, G, H, I
37 甲32×甲37	A, B, C, H, I
38 甲34×甲37	B, C, D, G, I

て  $T_2$  を特徴付け, 列ごとに1ノルムで正規化を行う.

そして, 以下の演算により, 20の印象語群とその重みを表すベクトル  $v_c$  を抽出する.

$$v_c = T_2 v_f \\ = (v_{c1}, v_{c2}, \dots, v_{c20}).$$

### 3.3 任意の言葉の印象に合致した言葉の音声表現の自動構成方式

本節では, 提案方式である, 任意の言葉の印象に合致した言葉の音声表現の自動構成方式について述べる. 本方式は, 3.2節に示した, 音相理論に基づいて任意の言葉からその音の印象に合致したメタデータを抽出する方式の逆演算として定義さ

$$a_k = \begin{cases} 1 & \text{if } x_i x_j \geq 1 \\ 0 & \text{if } x_i x_j < 1 \end{cases}$$

(5) 表情ベクトル  $v_f$  の表情語群ベクトル  $v_c$  への変換  
表4および5に示される78の表情と, それらから抽出される表情語群との関係を用いて変換行列  $T_2$  を作成する.  $T_2$  の行要素には表4および5に示された78の表情を, 列要素には表3に示された20の表情語群を配置し, 表4および5に示された表情と表情語群との関係を用いて, ある表情語を抽出する表情に対応する要素を“1”, それ以外の要素を“0”とすることによ

表6 初期ベクトルの構成に用いた要素群.

Table 6 Elements for the first vector.

ID	要素名	ID	要素名
00	ア音数	14	接近音数
01	イ音数	15	調音種数
02	ウ音数	16	R 数
03	エ音数	17	N 数
04	オ音数	18	Q 数
05	母音種数	19	濁音数
06	有声音数	20	H の総合音価
07	有声破裂音数	21	B の総合音価
08	無声破裂音数	22	高勁輝拍数
09	有声摩擦音数	23	逆接拍数
10	無声摩擦音数	24	拍数
11	無声摩擦音数	25	無声化母音数
12	鼻音数	26	無声拗音数
13	流音数	27	新子音数
		28	濁拗音数

表7 音相基の標準値.

Table 7 Typical values for elements.

音相基	標準値	音相基	標準値
ア音	32%	促音	1%(*)
イ音	20%	撥音	1%(*)
ウ・オ音	38%	濁音	10%(*)
エ音	16%	高勁輝拍	0.8%
有声音	53%	順接拍	77%
無声音	47%	逆接拍	23%
有声破裂音系	10%(*)	無声化母音	12%
無声破裂音系	10%(*)	無声拗音	8%
有声摩擦音	10%(*)	子音拍	28%
無声摩擦音系	10%(*)	新子音	12%
鼻音	10%(*)	濁拗音	12%
流音	10%(*)	摩擦音系	26%
長音	1%(*)		

(\*) 文献 [2] では値が明示されていないため、本稿では例語から類推した値を用いている。

れる。

### 3.3.1 任意の言葉と音相理論の表情語群との相関の計量

2 節で示した意味の数学モデルを用いて、入力された任意の言葉と表 3 に示されている A~T の 20 の表情語群それぞれの意味的な相関を計量することにより、これら 20 の表情語群それぞれの重みの大きさを表す 20 次元の表情語群ベクトル  $v_c$  を生成する。ここで、 $v_c$  の各要素の値は、入力された言葉と、その要素に対応する語群との相関の大きさである。例えば、 $v_c$  の 1 行目の要素は、表 3 の 1 行目の要素である表情語群 A に対応する表情語 "plain, obvious" と、入力された言葉との意味的な相関を意味の数学モデルにより計量した場合の相関の大きさとなる。

### 3.3.2 音相理論における「表情」の抽出

表情語群ベクトル  $v_c$  から、音相理論における甲類及び乙類の 78 の表情それぞれの相関を求める。ここでは、3.2 節 (5) における変換行列  $T_2$  の一般化逆行列  $T_2^\dagger$  を計算し、これを用いて  $v_c$  を甲乙それぞれの表情との相関からなる 78 次元のベクトル  $v_f$  に変換する。

$$v_f = T_2^\dagger v_c.$$

### 3.3.3 甲類表情のみの抽出

表 5 に示されるように、乙類表情は 2 つの甲類表情の論理積として表現することができる。逆演算により乙類表情が甲類表情に与える影響は無いと考えられるので、 $v_f$  を 40 種の甲類表情のみに対応する 40 次元のベクトル  $v_k$  に射影し、以下の手順ではこの  $v_k$  を用いることとする。

### 3.3.4 甲類表情の判定演算の逆演算

3.2 節の式 (1) における  $p_i$  は  $v_k$  の  $i$  番目の要素として表れてくるため、式 (1) を  $x_i$  について解いたものである

$$x_i = N(0.15p_i + n_i)$$

によって甲類表情の判定演算の逆演算を行い、式 (2) に示すような  $v_k'$  を生成する。

音相基	拍数	標準値
母音種比	1~2	判定対象外
	3	3 以上で「多用」、1 以下で「過少」
	4	3 以上で「多用」、1 以下で「過少」
	5~7	4 以上で「多用」、2 以下で「過少」
	8~10	「多用」は判定対象外、3 以下で「過少」
	11 以上	「多用」は判定対象外、4 以下で「過少」
調音種比	1	判定対象外
	2	3 以上で「高」、1 以下で「低」
	3~4	4 以上で「高」、2 以下で「低」
	5~6	5 以上で「高」、3 以下で「低」
	7~9	6 以上で「高」、4 以下で「低」
	10~14	7 以上で「高」、5 以下で「低」
	15 以上	「高」は判定対象外、7 以下で「低」
動性	1	H = 1.0
	2	H = 1.5
	3	H = 2.0
	4 以上	H = (拍数) × 0.6 - 0.1
輝性	1	B = ±0.1
	2	B = ±0.4
	3	B = ±0.6
	4 以上	B = ± (拍数) × 0.6 - 1.5
多拍	8 拍以上	
少拍	3 拍以下	

$$v_k' = (x_1, x_2, \dots, x_{40})^T. \quad (2)$$

なお、本稿では、標準値が段階的に変化する音相基に関しては  $x_i = 0$  として除外している。非線形な演算を含む音相基の抽出法は今後の課題とする。

### 3.3.5 音相基の抽出

3.2 節 (2) における変換行列  $T_1$  の一般化逆行列  $T_1^\dagger$  を用いることによって、以下のように  $v_k'$  から  $v_b$  への逆変換を定義することが可能となる。

$$v_b = T_1^\dagger v_k'.$$

### 3.3.6 出力される言葉の音の決定

本節では、 $v_b$  の各要素から、実際に出力される音の決定方法について示す。ここでは、出力される言葉の拍数が予め決定していることを前提としている。

#### (1) 母音数の決定

音相基ベクトル  $v_b$  から、出力される音がもつア音数、イ音数、(ウ音数 + オ音数)、エ音数、促音数および撥音数の割合が計算される。これら6つの音数の合計がその言葉の拍数に等しくなるように、相関の大きい順に音を採用していく。ただし、提案方式ではウ音とオ音の音数の比を決定することができないので、これらは確率0.5でランダムに配分するものとする。

#### (2) 子音数の決定

(1)と同様にして、子音に関しても音数を決定することができる。ここでは、有声破裂音数、無声摩擦音数、有声破擦音数、無声破擦音数、無声摩擦音数、鼻音数、流音数、接近音数が決定する。

#### (3) 制約条件による絞り込み

決定した母音種と子音種から、構成可能な全ての音の組み合わせを候補とする。そして、上記(1)及び(2)で利用されなかった音相基ベクトルの要素を制約条件とみなすことにより、候補の絞り込みを行う。

## 4. 実験例

提案方式の実現性及び有効性を検証するため、提案方式に基づく実験システムを構築して実験を行った。

### 4.1 実験環境

本実験では、意味の数学モデルを用いることによって、入力された任意の言葉と提案方式によるシステムとの間の連結を行った。意味の数学モデルの基本構成は2節に示す通りである。メタデータ空間  $MDS$  の設定については、“Longman Dictionary of Contemporary English” [8] という英英辞典を利用した。同辞書は約2,000語の基本語を用いて約56,000語の見出し語を説明している。この基本語を特徴とみなし、見出し語の説明で肯定的に使われている語を“1”、否定的に使われている語を“-1”、説明に使われていない語を“0”として2節(1)におけるデータ行列  $M$  を作成した。これにより、約2,000次元の正規直交行列であるメタデータ空間  $MDS$  が生成された。 $MDS$  では、約  $2^{2000}$  通りの意味の様相を表現可能である。

### 4.2 実験例

本実験では、任意の印象語をコンテキストとして入力することにより、提案方式による音声表現の自動構成を行う。ここでは実験例として、拍数4、コンテキストを1語とした場合の出力例を示す。

表8はコンテキストを“fun”とした場合の、表情語群の上位10件と出力された語の例である。この場合、D(動的)やM(適応性)、S(高尚)、N(清らかさ)といった表情語群の重みが大きくなり、kやtʃ、pなどの動性や輝性が比較的高く、かつ無声音である子音が多く使われていることが分かる。これは、“fun”というコンテキストに概ね合致していると考えられる。

表8 出力例 (context: fun) .

Table 8 experimental result(context: fun).

表情語群	相関量	出力語の例
D	0.434109	チャバテロ
M	0.412452	ラケクチャ
S	0.391166	ラボチャペ
N	0.385413	ケカチュラ
B	0.379831	テロタチャ
I	0.367643	ルタテチャ
E	0.364042	ルチャテバ
T	0.347459	カペラチュ
G	0.310286	ペチヨラカ

## 5. あとがき

本稿では、任意の印象に合致した言葉の音声表現の自動構成方式を提案した。提案方式により、音相理論に基づくメタデータ自動抽出方式の逆演算による、言葉の音声表現の自動生成が可能となる。

また、本稿では、入力された印象を音で表現する際に使われていると思われる音素の種類とその数を推測し、それらを用いた全ての組み合わせを候補とし、それらの中から、音相理論において特に重要な“逆接拍”や“無声化母音”など幾つかの要素のみに注目して候補を絞っていく方式を採用した。これにより、印象に合致した既存の言葉の検索や、自動構成の際に既存の言葉の変形による新しい音の生成などの機能の実現が可能となり、新しいネーミング手法やニックネームの提案など、新たな機能への応用が可能となると考えられる。

今後の課題は、言葉の意味や音声学などの知識を反映した音声の配置の実装や、既存の言葉やユーザが与えるテーマなどの言葉を基にした音声表現の構成方式の実現が挙げられる。

## 文 献

- [1] 木通隆行: “音相: 社名、商品名から人名までヒット・ネーミングは“音”で決まる”, プレジデント社, (1990).
- [2] 木通隆行: “日本語の音相 — ことばのイメージを捉える技術, 表現する技術 —”, 小学館スクウェア, (2004).
- [3] Homma, H., Nakanishi, T. and Kitagawa, T.: “A Method of Automatic Metadata Extraction Corresponding to the Impression by Sound of the Words”, The 16th European-Japanese Conference On Information Modelling and Knowledge Bases (EJC2006), 2006
- [4] 本間秀典, 中西崇文, 北川高嗣: “任意の言葉を対象とした音の印象によるメタデータ自動抽出方式”, 電子情報通信学会技術研究報告, Vol.106, No.149, pp.7-12, 電子情報通信学会 (2006).
- [5] Kitagawa, T. and Kiyoki, Y.: “The mathematical model of meaning and its application to multidatabase systems”, Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems, pp. 130-135(1993).
- [6] Kiyoki, Y., Kitagawa, T. and Hayama, T.: “A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning”, Multimedia Data Management – using metadata to integrate and apply digital media –, McGrawHill, A. Sheth and W. Klas(editors), Chapter 7 (1998).
- [7] 清木康, 金子昌史, 北川高嗣: “意味の数学モデルによる画像データベース探索方式とその学習機構”, 電子情報通信学会論文誌, D-II, Vol. J79-D-II, No. 4, pp. 509-519 (1996).
- [8] Longman Dictionary of Contemporary English, Longman (1987).