

任意の言葉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[†] 中西 崇文^{††} 北川 高嗣[†]

[†] 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 情報数理研究室

^{††} 情報通信研究機構 〒619-0289 京都府相楽郡精華町光台 3-5

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持った言葉の自動生成方式を提案する。これまで、音相理論と呼ばれる理論を用いて、入力された任意の言葉から、その音声表現から人間が受ける印象を表す言葉をメタデータとして自動抽出する方式を実現している。本稿では、音相理論における音声表現が持つ音相基と呼ばれる要素と、それらから抽出される印象語群の間の相関を求めることにより、メタデータ自動抽出方式の逆演算を行い、言葉を自動構成する方式を提案する。さらに、本方式を意味の数学モデルに適用することによって、入力された任意の言葉と印象語群の相関を計量し、任意の言葉による、その印象に合致した音声表現の自動構成方式の実現可能性について述べる。

キーワード 意味の数学モデル, 意味的連想検索, 逆演算, 音相, 感性

An Automatic Composition Method of Sound Expression of a Word Corresponding to the Impressions of Arbitrary Words

Hidenori HOMMA[†], Takafumi NAKANISHI^{††}, and Takashi KITAGAWA[†]

[†] Graduate School of Systems and Information Engineering, University of Tsukuba

^{††} National Institute of Information and Communications Technology

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 of “Onso”. In this paper, we present an inverse operation of the automatic metadata extraction method. This operation enables an automatic composition of the sound expression of a word.

Key words Mathematical model of meaning, Semantic associative search, Inverse operation, Onso, Kansei

1. まえがき

コンピュータネットワークの広域化・高速化により、コンピュータを介して様々な情報群と接触する機会が増大している。このような環境において、多種多様なメディアデータがネットワーク上に散在しつつあり、これらを対象とした、高度な情報検索方式、及び知識の獲得方式が求められている。

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が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。提案方式がユーザとシステムとの間の情報伝達を

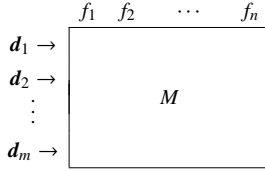


図1 データ行列 M によるメタデータの表現.
Fig.1 Representation of the data matrix M .

可能とする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、人間の直感や感性により合致したインターフェースの実現が可能となると考えられる。

これまで、文献[5]~[7]で、言葉と言葉の関係の計量による検索機構として、意味の数学モデルを提案している。これは、単語群を文脈として解釈する機構により、言葉と言葉、あるいは、言葉と検索対象のメディアデータ、ドキュメント間を文脈に応じて動的に計算することを可能とする。意味の数学モデルでは、検索対象をベクトル化し、メタデータ空間と呼ばれる空間に写像する。さらに、それらのベクトルをメタデータ空間の部分空間に射影して計量することにより、文脈に応じた連想検索を実現している。

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[5]~[7]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的なコミュニケーションメディアの実現の第一歩となりうる。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

本節では、言葉と言葉の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。

具体的には以下の通りメタデータ空間 MDS を設定する。

初めに、 m 個の基本データについて各々 n 個の特徴 (f_1, f_2, \dots, f_n) を列挙した特徴付ベクトル $d_i (i = 1, \dots, m)$ が与えられているものとし、そのベクトルを並べて構成する $m \times n$ 行列を M とおく(図1)。このとき、 M は、列ごとに2ノルムで正規化されている。

- (a) データ行列 M の相関行列 $M^T M$ を計算する。
- (b) $M^T M$ を固有値分解する。

$$M^T M = Q \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_v & \\ & & & 0_{\dots 0} \end{pmatrix} Q^T,$$

$$0 \leq v \leq n.$$

ここで行列 Q は、

$$Q = (q_1, q_2, \dots, q_n)$$

である。この $q_i (i = 1, \dots, n)$ は、相関行列の正規化された固有ベクトルである。相関行列の対称性から、この固有値は全て実数であり、その固有ベクトルは互いに直交している。

- (c) メタデータ空間 MDS を以下で定義する。

非ゼロ固有値に対応する固有ベクトルによって形成される正規直交空間をメタデータ空間 MDS と定義する。この空間の次元 v は、データ行列 M のランクに一致する。この空間は、 v 次元ユークリッド空間となる。

$$MDS := span(q_1, q_2, \dots, q_v).$$

$\{q_1, \dots, q_v\}$ は MDS の正規直交基底である。

(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

設定されたメタデータ空間 MDS へ、メディアデータのメタデータをベクトル化し写像する。これにより、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。

具体的には、メディアデータのメタデータを以下のようにベクトル化する。

- (a) メディアデータの特徴づけ

メディアデータ P を t 個の印象語(あるいは、 t 個のオブジェクト) o_1, o_2, \dots, o_t を用いて、次のように特徴づける。

$$P = \{o_1, o_2, \dots, o_t\}.$$

ここで、各印象語 o_i は、データ行列の特徴と同一の特徴を用いて表現される特徴付ベクトルである。

$$o_i = (o_{i1}, o_{i2}, \dots, o_{in}).$$

- (b) メディアデータ P のベクトル表現

メディアデータ P を構成する t 個の印象語 o_1, o_2, \dots, o_t が、それぞれ n 次元のベクトルで定義されている。印象語 o_1, o_2, \dots, o_t は、合成することで n 次元ベクトル表現され、メディアデータベクトル p を形成する。さらにこのメディアデータベクトル p をメタデータ空間 MDS に写像する。これにより、同じ空間上に言葉とメディアデータが配置されることになり、言葉とメディアデータの間を空間上の距離として動的に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。このコンテキストを用いてメタデータ空間 MDS に各コンテキストに対応するベクトルを写像する。これらのベクトルは、メタデータ空間 MDS において合成され、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、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ときに用いられる器官 (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「勁性」(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, 勁性の関係は文献 [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、数値が大きいくほど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。

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 +B1.7 H1.6	t +B1.4 H1.5	k +B1.3 H1.3
	有	b -B2.2 H1.2	d -B2.0 H1.0	g -B2.0 H1.0
破擦音	無		ts +B1.7 H2.2	tʃ +2.2 H2.7
	有		dʒ -B2.0 H1.0	çʒ -B1.5 H1.5
摩擦音	無	f B0.0 H0.0	s +B0.3 H0.7	ʃ +B0.8 H1.2
	有			h B0.0 H0.0
弾き音 (流音)	有		r +B0.4 H0.8	
	有	m B0.0 H0.0	n B0.0 H0.0	ŋ B0.0 H0.0
接近音	有	w -B0.5 H0.0	j +B0.5 H0.5	

- ・ 無：無声音
- ・ ts：ツの子音
- ・ dʒ：ジを除くザ行の子音
- ・ f：フおよびファ行の子音
- ・ ŋ：ガ行鼻音
- ・ 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
- ・ 有：有声音
- ・ tʃ：チおよびチャ行の子音
- ・ çʒ：ジ (ヂ) およびジャ行の子音
- ・ ʃ：シおよびシャ行の子音
- ・ j：ヤ行子音および拗音の第 2 子音

表 2 音素と音価表 (母音の場合).

Table 2 The table of Onso for vowels.

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

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

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, 澁刺さ	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

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,J,L,O
19. N(撥音) 多用	Q
20. 濁音多用	P,Q,R,S,T
21. 高勁性	A,B,C,D,E,G,H,J,L
22. 低勁性	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,J,K,M
25. マイナス輝性	Q,R,S
26. マイナス高輝性	Q,R,T
27. 高勁輝拍多用	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

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.

$$v'_k = T_1 v_b \\ = (x_1, x_2, \dots, x_{40})^T.$$

(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 ± 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.

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

(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

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”とすることによって 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。

$$v_c = T_2 v_f$$

$$= (v_{c1}, v_{c2}, \dots, v_{c40}).$$

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	有声破擦音数	23	逆接拍数
10	無声破擦音数	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

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れる。

3.3.1 任意の言葉と音相理論における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と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の表情それぞれとの相関を求める。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。計算上、乙類表情が逆に甲類表情に与える影響は無いと考えられるので、入力された言葉と甲類表情との相関が抽出できれば、音相基ベクトル v_b を抽出することが可能となると考えられる。従って、ここでは、 v_f を40

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。

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」
勁性	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	

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。

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

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、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それぞれの音との相関が得られる。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) 制約条件による音の絞込み

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は音を決定するための制約条件となる。ここでは、相関の大きさに従って各音相基の数を決定し、決定したそれぞれの音相基に合致するように母音と子音の組み合わせや並びを決定していく。

(4) 出力される音の決定

(3)までで決定しなかった音素を、(1)及び(2)の結果を用いて決定していく。複数の音素が割り当てられている要素に関しては、等確率でランダムに選択されることとする。

4. 実験

提案方式の実現可能性を検証するため、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

表 8 出力例 (context: light) .

Table 8 experimental result(context: light).

表情語群	相関量	出力語の例
T	0.480804	ムチャピン
E	0.428942	キンチャヌ
D	0.412870	ピチャンヌ
M	0.412719	チョマピン
S	0.362957	チュンニタ
I	0.358367	チョンタニ
B	0.326667	チャミント
F	0.311155	タンニツ
N	0.290043	チンプマ

2^{2000} 通りの意味の様相を表現可能である.

4.2 実験方法

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, 構成する音の拍数としてある一定の値を予め入力しておくこととする.

4.3 実験例

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の例を示す.

表 8 はコンテキストを “light” 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. 表情語群の上位は E (派手さ) や D (動的) といった表情が多く抽出されている反面, T (静的) が抽出されているため, k や tʃ, p など明るいイメージを持つ子音に混じって m や ŋ などの音が含まれている.

表 9 はコンテキストを “happy” 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. こちらでは M (適応性) や D (動的), S (高尚), E (派手さ) などの表情が上位に抽出され, “light” の場合よりもより明るく軽快なイメージを与える子音が多く含まれていると考えられる.

表 10 は, コンテキストが “dull” の場合の実験結果の例である. ここでは, T (静的), S (高尚), M (適応性), D (動的), E (派手さ) など様々な表情が抽出された結果, 明るい子音である k や, 暗い子音である ɸ, マイナスの輝性を持つ新子音 w など, 様々な子音が含まれた言葉が出力された. “light” や “happy” の出力例に比べ暗い印象を持つ子音が選択されやすいことから, この結果は, “dull” の持つやや重いイメージに合致すると考えられる.

これらの結果から, 提案方式による, 任意の言葉に合致する音声表現を持つ言葉の自動構成の実現可能性が示された. しかしながら, 例えば表 8 の “キンチャヌ” と表 10 の “キンチャユ” のように, 違う印象語を用いた場合で似通った言葉が出力されるケースも存在した. これは, 音声表現の決定を音素ごと, つまり子音と母音を完全に分けて行ったために起こったと考えられる. 出力される言葉が持つそれぞれの拍が決定した後の, 全体のイメージをよりの確に反映できるような制約条件の追加や改善は今後の課題である.

表 9 出力例 (context: happy) .

Table 9 experimental result(context: happy).

表情語群	相関量	出力語の例
M	0.428685	カヒミン
D	0.399780	ピンシマ
S	0.398834	ニンキファ
E	0.387301	ヒンミパ
T	0.384043	キナンシ
I	0.377768	ニシャキン
K	0.354037	ニヒンタ
B	0.349129	ナキシン
N	0.347991	ピンサニ
A	0.290812	ミキサン

表 10 出力例 (context: dull) .

Table 10 experimental results(context: dull).

表情語群	相関量	出力語の例
T	0.517440	クジチャン
S	0.503980	キンチャユ
M	0.460738	ウインタチュ
D	0.409198	ヤチョンピ
E	0.373566	ウィチャンボ
I	0.319162	カチュンジ
B	0.309764	チョウイタン
G	0.308489	ワンチョコキ
A	0.303535	キンワチュ
C	0.301518	ウオンタチ

5. あとがき

本稿では, 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した. 提案方式により, 音相理論に基づく言葉の音の印象に合致したメタデータの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方式の実現可能性が示された.

また, 本稿では最小の制約条件のみを用いて音声表現の自動構成方式を実現した. このことにより, ユーザの入力した言葉の印象に合致した, 存在しない言葉を含む様々な音声表現を自動構成することが可能となった. 提案方式において出力される音相基ベクトル v_b から音声表現を決定する際に, 音声学や言語学, 自然言語などにおける知識を新たな制約条件として導入することによって, 新しい言葉だけでなく, より発音しやすい言葉や, ユーザの要求に合致した印象と意味を兼ね備えた言葉を限定して出力させるようにするなどの機能の実現が可能となる.

さらに, 提案方式を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 [3], [4] と組み合わせることにより, 元となる言葉の音声を, 入力された言葉の印象により近いものとなるように変更していくような操作も実現可能になると考えられる. これにより, 提案方式による, 任意の言葉の印象に合致した新しいネーミング手法の提案やニックネームの提案など, 新たな機能の実現が可能となると考えられる.

しかしながら, 実験結果から明らかなように, おおまかな印象を捉えてはいるが, 全く無意味な言葉が出力される傾向が強

いことが確認された。これは、本稿ではまだ一部の制約条件を利用していないことや、音声の配置の大部分を乱数により決定していることが影響していると考えられる。今後の課題として、音相基の非線形な要素に対応する制約条件の実装や、言葉の意味や音声学などの知識を反映した、音声の配置に関する制約条件の改善などが挙げられる。

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).