情報処理学会論文誌 Vol.63 No.4 1159–1169 (Apr. 2022)

自然で表現豊かな笑い声合成に向けた感情情報からの笑い声の構成要素決定法

有本 泰子[1,a)] 今西 利於[1,†1] 森 大毅[2]

受付日 2021年8月24日, 採録日 2022年1月11日

概要：自然で表現豊かな笑い声合成の実現のために，その入力情報となる笑い声の構成要素を決定する方法を提案する．従来手法では，笑い声を合成する際に入力する笑い声の構成要素の並びを人手により指定する必要があった．また，入力とする笑い声の構成要素を記述するためには音響音声学的知識が不可欠であるため，ユーザが直感的に入力情報を記述することは困難である．本研究では，より直感的な操作による自然で表現豊かな笑い声合成の実現に向けて，入力情報となる笑い声の構成要素の並びをその感情情報を利用して決定する手法について提案する．まずは，構成要素と感情情報との関係を明らかにするため，笑い声を構成する無声呼気・有声呼気・無声吸気・有声吸気の各音響イベント間で聴取実験により付与された感情3次元（快–不快，覚醒–睡眠，支配–服従）に差があるか検証した．その結果，いずれの感情次元においても，構成要素間に有意な差が認められた．さらに，提案手法を用いて入力となる構成要素列を決定した笑い声を合成し，その自然性と感情知覚を評価する聴取実験を行った．その結果，提案手法を用いて笑い声の構成要素を自動決定しても，自然性を損なうことなく，感情知覚させることのできる笑い声の合成が可能であることを示した．

キーワード：笑い声合成，感情次元，笑い声の構成要素，ディリクレ回帰，WaveNet

Laughter Components Estimation Using Emotional Information towards Natural and Expressive Laughter Synthesis

Yoshiko Arimoto[1,a)] Rio Imanishi[1,†1] Hiroki Mori[2]

Received: August 24, 2021, Accepted: January 11, 2022

Abstract: Towards natural and expressive laughter synthesis, a determination method of the laughter components is proposed for automatically creating input information for laughter synthesis system. The conventional laughter synthesis requires manual arrangement of laughter components when synthesizing laughter. In addition, it is difficult for end-users to intuitively arrange them to synthesize laughters because acoustic and phonetic knowledge should be required to arrange its complicated structure of laughter components. To synthesize natural and expressive laughters by more intuitive operation, this study proposes a method to determine the sequence of laughter components as the input information using the emotional information. First, to clarify the relationship between each laughter component (unvoiced exhalation, voiced exhalation, unvoiced inhalation, and voiced inhalation) and emotional information, statistical tests were performed on the factor of laughter component using the perception scores on three emotional dimensions (pleasantness, arousal, and dominance). The result revealed the significant differences between the four components for all emotional dimensions. Furthermore, a listening test was conducted to evaluate the naturalness and emotional perception of synthesizing laughter using the input component sequence determined by the proposed method. As a result, it was found that the synthesized laughter using our approach is emotionally perceivable without losing its naturalness.

Keywords: laughter synthesis, emotion dimension, laughter component, Dirichlet regression, WaveNet

1. はじめに

ユーザの要望を音声で認識し，リクエストに応える対話型のインタフェースである音声アシスタントやバーチャルエージェントなどの音声合成技術を用いた対話システ

[1] 千葉工業大学
Chiba Institute of Technology, Narashino, Chiba 257–0016, Japan
[2] 宇都宮大学
Utsunomiya University, Utsunomiya, Tochigi 321–8585, Japan
[†1] 現在，株式会社エーアイ
Presently with AI, Inc.
[a)] ar@mac-lab.org

ムが広く親しまれている．最新の Text-to-Speech（TTS）による音声合成技術は非常に高いレベルに達しており，Sequence-to-Sequence 型のニューラルネットによってテキストから推定されたメルスペクトログラムから WaveNet [2] を利用して音声波形を生成する Tacotron2 [1] は人間が発する発話と聞き分けができないほど自然性が高い合成音声を実現可能である．そうしたなかで，音声合成技術には音声の音質を向上させることだけでなく，喜び，怒り，悲しみなどの感情を表現した発話音声を実現することも求められてきている [3], [4]．一方で，感情伝達やソーシャルインタラクションを構成する要素である笑い声 [5], [6] を合成することはあまり検討されていない．Mori ら [7] が WaveNet を用いた笑い声合成を提案しているが，人間の発する笑い声に比べ自然性が低いことが示されており，自然な笑い声合成の実現はいまだ困難な課題であることが分かる．

また，TTS を用いた笑い声合成には入力情報に用いる記号列に関する問題がある．TTS を用いた発話音声の合成では入力情報に生成したい言語的内容を示した記号列であるテキストを用いることで発話を合成するが，笑い声では発話の言語的内容に相当する情報を記述する記号が標準化されておらず，笑い声合成を行う研究ごとに異なる記号列を用いているのが現状である [7], [8]．たとえば，Mori らは笑い声を構成する要素である bout や吸気音を入力情報とし [7]，Tits らは有声/無声の記号列を入力情報としている [8]．こうした bout や吸気音，有声/無声といった笑い声を構成する要素は，合成する笑い声に社会的シグナルとしての特徴を保持させるために非常に重要な情報と考えられる [9], [10]．一方で，これらの構成要素の理解には高度な音声学・音響学の知識を必要とするため，笑い声合成器に与える構成要素列を，笑い声を合成したい一般のユーザが操作することは非常に困難な作業である．

一方，笑い声が感情次元や態度などの非言語的・パラ言語的情報を伝達する社会的なシグナルであることが明らかになっている [5], [6]．笑い声と感情情報や印象に関する研究の多くは，笑い声の種類を識別/分類するものであり [11], [12]，感情次元情報を用いてカテゴリ識別を行う研究 [11] や笑い声の構成要素と感情次元の関係を分析する研究 [12] が行われている．また，笑い声を構成する要素が異なると伝達される情報も異なることが示されている [5], [9], [12]．笑い声の構成要素と感情次元の関係を分析する研究では，笑い声の構成要素によって感情次元知覚特性が異なることが明らかになっている [12]．つまり，笑い声の構成要素列を変化させることで，表現豊かな笑い声の合成の実現が可能であることが示唆されるとともに，感情情報から笑い声の構成要素列を推定することができる可能性がある．

笑い声の構成要素を感情情報から決定し，隠れマルコフモデル（Hidden Markov Model: HMM）による音声合成の入力情報として利用しても，合成した笑い声の自然性は損なわれないことが明らかになっている [13]．しかし，この際に利用された感情情報は感情強度（emotional intensity）のみであり，心理学的な感情次元モデルである快–不快・覚醒–睡眠次元 [14] や支配–服従次元 [15] は利用されていない．Urbain らが文献 [13] の研究で感情強度のみを対象として笑い声合成を実現可能としている背景には，対象としている笑い声を陽気な笑い声（hilarious laughter）に限定していることがあげられる．合成の対象を陽気な笑い声に限定することによって，個々の笑い声の違いを表現可能な感情次元は感情強度だけとなる．一方で，笑い声には陽気な笑いだけでなく，嘲笑や失笑などの多様な表現が存在する．笑い声の種類を限定せずに多様な表現を可能とするために，快–不快・覚醒–睡眠・支配–服従次元などの多次元を利用した構成要素推定が求められる．

本研究では笑い声合成の入力情報に心理学的な感情次元モデルである快–不快・覚醒–睡眠・支配–服従の 3 次元を用いることで，笑い声の構成要素を推定し，自然性を損なわずに表現豊かな笑い声を合成することを目的とする．これにより，音声学や音響学の知識がないユーザが感情情報を直感的に操作することで自然で表現豊かな笑い声を合成することが可能となる．たとえば，筋萎縮性側索硬化症（ALS）などで気管切開や喉頭摘出を行った人が，音声合成器で笑い声を生成したいと思ったときに，音声学や音響学の知識を要する複雑な構成要素列を入力することなく，そのときに表現したい自身の感情状態を入力することで，多様な表現の笑い声を合成することが可能となる．

本研究では，先行研究よりもより柔軟な構成要素列を作成するために，まずは複数の呼気から構成される bout を，call と呼ばれる 1 つ 1 つの呼気による音響イベントに分割し，それぞれを有声/無声の 2 種類に分類する call ラベリングを行って，それぞれの構成要素と感情情報との関係を調べた．次に，感情情報から 1 つの笑い声を構成する各要素の割合を推定するために，ディリクレ回帰を用いた構成要素推定モデルを作成した．さらに，構成要素推定モデルにより推定した各要素の割合を参照して構成要素列を決定し，実際に笑い声合成を行った．提案手法の有効性を評価するため，従来手法と提案手法で合成した笑い声に対し，自然性と感情次元知覚を評価する聴取実験を行い，結果を比較した．

2. 関連研究

2.1 笑い声の構成要素に関する研究

笑い声の構成要素に関する研究はいくつか行われている [7], [16]．Truong らは笑い声アノテーションの基本的な枠組みを定義し，異なるシチュエーションで発せられた笑い声に対し，アノテーションを行っている [16]．この研究では笑い声は episode（epi）と定義され，episode には連続

した呼気音を示す bout，単一の吸気音を示す onset，offset，笑い声の特徴を持った発話を示す speech-laugh（spl），笑顔に聞き取れる発話を示す smiled speech（smsp）から構成されている．これらの各要素は 1 つまたは複数によって episode を構成している．そのなかでも，bout は有声音（vcd），無声音（uvd）に区別されている．また，bout 内には音節に相当するパルスが複数存在し，それぞれのパルスが有声音である場合は vp，無声音である場合は up と定義されている．森らは笑い声の構成要素を笑い声合成に利用するため，自発対話コーパスに対してアノテーションを行っている [17]．この研究では笑い声を laughter episode（laugh）と定義している．Truong らとは異なり，laughter episode は bout，声帯振動をともなわない吸気音である無声吸気（h），声帯振動をともなう吸気音である有声吸気（H）から構成されている．これらの各要素は 1 つまたは複数によって laughter episode を構成している．

2.2 笑い声合成に関する研究

笑い声合成に関する研究はいくつか行われている [7], [8], [18]．Nagata らは context を考慮した HMM 音声合成による笑い声合成を行っている [18]．context には call の音韻，先行・後続セグメントの音韻，発話における bout の位置，bout における call の位置，bout を構成する call の数があげられている．これらの context を考慮することで生成される笑い声の自然性を向上させることができることを示している．Mori らは WaveNet を用いた笑い声合成を行っており，HMM 音声合成に比べ WaveNet の方が自然性が向上することを示した [7]．Tits らは学習に用いる笑い声の少なさを補うため，発話による学習を活用した転移学習を利用した Seq2Seq による笑い声合成を提案している [8]．この手法によって生成される笑い声は HMM 音声合成によって生成される笑い声と同等の自然性を示している．また，ニューラルボコーダに自己回帰を行わず並列計算が可能な MelGAN [19] を用いることで自然性をより向上させることができることも明らかにしている．

2.3 笑い声の印象評価に関する研究

笑い声の印象評価に関する研究では，Bachorowski らが有声の笑い声と無声の笑い声に対し，友好的かどうかや親しみやすさなどの評価をさせている [9]．その結果，有声の笑い声と無声の笑い声で印象が変わることが明らかとなった．また，男性の笑い声と女性の笑い声間でも印象に差があることが明らかとなっている．Szameitat らは笑い声を 4 つのカテゴリ（joy，tickling，traunting，schadenfreude）に分類し，4 つの感情次元（覚醒–睡眠，支配–服従，話者の快–不快，聞き手に向けた話者の快–不快）の評価を行っている [11]．判別分析によって感情次元評価値から 73.1%の正解率で笑い声を分類できることを示し，感情次元とカテゴリの関係性が強いことを示した．また，感情次元評価値と音響的特徴量との関係を分析しており，発話における感情次元と音響的特徴量との関係と類似している点があることが明らかとなっている [20]．小山らは笑い声の表現の種類を表すカテゴリ（陽気な笑い，照れ笑い，愛想笑い，苦笑い，悲観的な笑い，嘲笑い，高圧的な笑い）ごとの笑い声の構成要素（bout，無声吸気，有声吸気）の割合を示しており，カテゴリによって構成要素の含まれる割合が異なることが明らかにしている [10]．

3. 音声資料

3.1 笑い声資料

本研究では，先行研究 [7], [12], [17] によりすでに笑い声ラベルと感情次元値が付与されている感情評定値付きオンラインゲーム音声チャットコーパス（OGVC）[21] を使用する．OGVC にはオンラインゲームをプレイ中の男性 9 名，女性 4 名の計 13 名のプレイヤの音声を収録した自発対話音声 9,114 発話が収録されている．このうち男性 5 名，女性 2 名の音声に対して，ひと続きの笑い声である laughter episode ラベル（{laugh}）と laughter episode を構成する 3 つの構成要素のラベル（一連の呼気に対する bout ラベル（b），無声吸気ラベル（h），有声吸気ラベル（H））が付与されている．7 名分の話者のデータには 1,381 個の {laugh} のラベルが付与されており，それぞれの構成要素の個数は bout が 1,650 個，無声吸気が 1,308 個，有声吸気が 797 個である．laughter episode の平均継続時間長は 0.95 秒（最小値は 0.07 秒，最大値は 8.30 秒）であった．

笑い声に付与されている感情次元は Russell [14] が提唱している快–不快，覚醒–睡眠と Mehrabian ら [15] が提唱している支配–服従の 3 次元である．付与された感情次元値は −3 から 3 の実数値をとり，値が低いと不快・睡眠・服従の状態を示し，値が高いと快・覚醒・支配の状態を示す．

3.2 call ラベリング

笑い声の構成要素をより柔軟に操作するため，複数の音響イベントから構成される bout に対して call ラベリングを行い，声帯振動をともなわない無声呼気には n を，声帯振動をともなう有声呼気には v のラベルを付与した．作業手順を以下の (1) から (3) に示す．

(1) bout の音声の聴取を行い，瞬間的な音の区切れや音の強弱などから bout を構成している call の個数を決定する．
(2) bout の波形とスペクトログラムの視察により，(1) で決定した call の個数をもとに各 call の時間的な区切りを付与する．
(3) 時間的に区切られた call に対し，無声呼気と判断した場合は n と有声呼気と判断した場合は v と記述する．

以上の作業手順に基づいて笑い声のラベルが付与されて

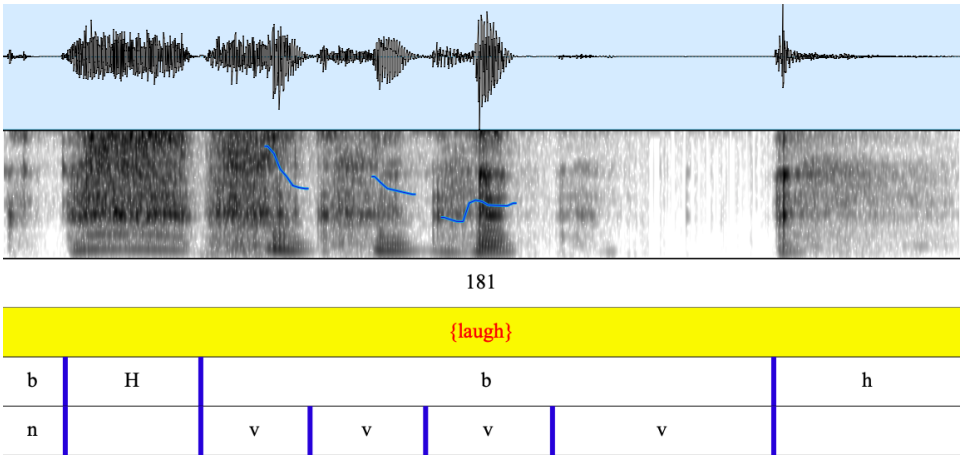

図 1 call のラベルの例（1 段目：音声波形，2 段目：スペクトログラム，3 段目：発話番号，4 段目：発話内容の転記（{laugh} は laughter episode を示す），5 段目：構成要素ラベル（bout（b），有声吸気（H），無声吸気（h）），6 段目：call ラベル（有声呼気（v），無声吸気（n）））

Fig. 1 Example of call labelling (1st layer: speech waveform, 2nd layer: spectrogram, 3rd layer: speech ID, 4th layer: transcription ({laugh} indicates laughter episode), 5th layer: laughter component labeling (bout (b), voiced inhalation (H), unvoiced inhalation (h)), 6th layer: call labeling (voiced exhalation (v), unvoiced exhalation (n))).

いる男性 5 名，女性 2 名のデータに対して call ラベリングを男性 1 名（ラベラ A）が行った．Call ラベリングは音声アノテーションツールである Praat [22] を用いて行った．**図 1** に call ラベリングを行った例を示す．図 1 中の 1 段目には音声信号が，2 段目にはそのスペクトログラムが表示されている．3 段目には発話番号が記されている．4 段目には発話内容の転記が示されており，図 1 では，laughter episode を示すラベルである {laugh} が記載されている．4 段目に {laugh} が記載された場合，5 段目と 6 段目にそれぞれ構成要素ラベルと call ラベルが付与される．5 段目の構成要素ラベルでは，bout には b を有声吸気には H を無声吸気には h が付与されている．構成要素ラベルとして bout を示す b が付与された箇所に対し，bout を構成する有声呼気（v）および無声呼気（n）の 2 種類の call ラベルが 6 段目に付与されている．

ラベリング作業によって付与された各構成要素の個数は，無声呼気が 1,308 個，有声呼気が 2,430 個である．ラベラ A の付与したラベルの妥当性を確認するため，ラベラ A 以外の男性 2 名のラベラ（B および C）に女性話者 1 名（06_FWA）の音声ファイルにおける冒頭 30 分間にラベリングを行わせ，一致度（Cohen's Kappa）を求めた．その結果，kappa 係数は 0.60 以上となり（それぞれ 0.68，0.70），上記の基準によるラベラ A のラベリングが妥当であることが示された．また，call ラベルの継続時間長がラベラ A と他 2 名のラベラ B および C とどのくらい差があるか確認するため，二乗平均誤差（RMSE）を 2 名の男性それぞれに求めた．その結果，v では RMSE が 0.13 秒（A vs. B）と 0.11 秒（A vs. C），n では 0.18 秒（A vs. B）と 0.17 秒（A vs. C）であった．

表 1 各構成要素の平均継続長と標準偏差

Table 1 The mean and standard deviation of duration for each component.

Components	Duration	σ
n	0.20	0.12
v	0.20	0.12
h	0.28	0.15
H	0.26	0.14

4. 笑い声の構成要素と感情次元の関係

4.1 分析目的

感情情報から 3.2 節でラベリングを行った各構成要素を決定するためには，感情情報と各構成要素との関係を明らかにする必要がある．ここでは，無声呼気，有声呼気，無声吸気，有声吸気の各構成要素に対する感情知覚特性の違いを分析する．

4.2 分析方法

3.1 節で説明した笑い声資料では，感情次元値は laughter episode 単位で付与されており，その構成要素 1 つ 1 つには感情次元値が付与されていない．しかし，各構成要素の平均的な長さは 0.20〜0.28（標準偏差は 0.12〜0.15）と非常に短いため（**表 1**），その短い区間の情報呈示のみだけで構成要素の感情次元を人間が評価することは困難である．そのため，各構成要素にはその構成要素が含まれてい

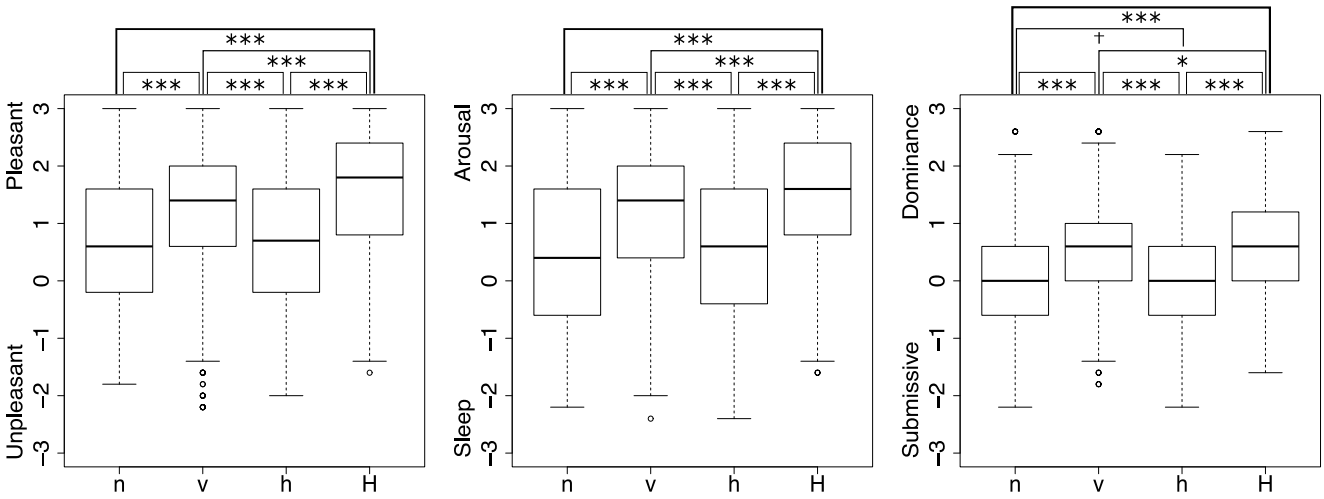

図 2 各感情次元における笑い声の構成要素の分布（左：快–不快，中央：覚醒–睡眠，右：支配–服従，$\dagger p < 0.1$，$* p < 0.05$，$*** p < 0.001$））

Fig. 2 Distribution of laughter components for each emotion dimension (left panel: Pleasant–Unpleasant, center panel: Aroused–Sleepy, right panel: Dominance–Submissive, $\dagger p < 0.1$, $* p < 0.05$, $*** p < 0.001$).

る laughter episode の感情次元値を割り振ることとした．割り当てられた感情次元値（快–不快，覚醒–睡眠，支配–服従）に対し，笑い声の構成要素の種類（無声呼気，有声呼気，無声吸気，有声吸気）を要因とした一元配置分散分析と Tukey-Kramer の HSD 検定を行った．

4.3 結果と考察

図 2 に各構成要素の感情次元値の分布を感情次元ごとに箱ひげ図で示す．一元配置分散分析の結果，快–不快が $F(3,5{,}840) = 196.73$，$p < 0.001$，覚醒–睡眠が $F(3,5{,}840) = 216.69$，$p < 0.001$，支配–服従が $F(3,5{,}840) = 205.72$，$p < 0.001$ となりすべての次元で構成要素の種類の主効果が有意であった．多重比較の結果，有声吸気（H）が各感情次元において最も高い感情評価値を示し，他の構成要素より快，覚醒，支配寄りに有意に高いことが明らかになった（H：平均値 1.56，標準偏差 1.00（快–不快），平均値 1.51，標準偏差 1.04（覚醒–睡眠），平均値 0.63，標準偏差 0.88（支配–服従），$p < 0.001$ または $p < 0.05$）．次に高い値を示したのは有声呼気（v）であり，無声呼気（n），無声吸気（h）より快，覚醒，支配寄りに有意に高いことが示された（v：平均値 1.30，標準偏差 1.04（快–不快），平均値 1.22，標準偏差 1.08（覚醒–睡眠），平均値 0.53，標準偏差 0.80（支配–服従），$p < 0.001$）．さらに，無声吸気（h）が無声呼気（n）より支配寄りに有意に高い傾向を示した（h：平均値 0.05，標準偏差 0.88，n：平均値 −0.04，標準偏差 0.88，$p < 0.1$）．

以上の結果から，笑い声の構成要素ごとに知覚させる感情が異なることが明らかとなった．これらの要素を組み合わせることによって，多様な感情を伝える笑い声を作り出すことが可能であることが示唆される（構成要素の組合せと感情次元の関係性については次の 5.1 節で検証する）．逆にこの構成要素の感情知覚特性を利用すれば，作成したい笑い声の表現がどのような感情であるかを指定することによって，その感情を表現することのできる笑い声の構成を決定することも可能である．つまり，笑い声を合成する際，入力情報として感情 3 次元を利用し各構成要素を含める割合を変えることで聞き手に知覚させる感情を操作することができる可能性がある．

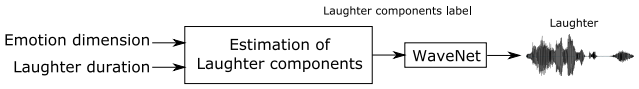

図 3 笑い声合成システムの構成図

Fig. 3 Structure of the laughter synthesis system.

5. 笑い声の構成要素推定モデル

5.1 構成要素の割合推定モデル

本研究では笑い声合成の合成時の入力情報に用いられている笑い声の構成要素を感情情報から推定するモデルを作成した．**図 3** は笑い声合成システムの構成図である．合成器は膨張たたみ込みの積層に基づく自己回帰ニューラル波形モデルである WaveNet [2] を用いている．WaveNet の入力には笑い声の構成要素（無声呼気，有声呼気，無声吸気，有声吸気）列を指定する．笑い声の構成要素列に含まれる各構成要素の割合は，生成したい笑い声の感情次元値（快–不快，覚醒–睡眠，支配–服従）を用いたディリクレ回帰によって推定する．推定した笑い声の構成要素の割合からラベルの作成を行い，作成したラベルを用いて笑い声を作成する．

1 つの笑い声を構成する各要素の割合を感情情報から推定するためにディリクレ回帰モデルを作成する．ディリク

表 2 モデルの各回帰係数と切片
Table 2 Regression coefficients and intercepts for our model.

Components	β_1	β_2	β_3	β_4	β_0
n	−0.04	0	−0.20	0.51	−2.06
v	0.22	0.16	0.23	0.22	−1.60
h	0.07	−0.10	−0.18	0.49	−1.92
H	0.09	0.19	−0.04	0.36	−2.24

表 3 構成要素の割合推定モデルの推定精度（RMSE）
Table 3 Accuracy (RMSE) of the estimated laughter component ratio.

Components	RMSE (%)	RMSE (s)
n	0.2899	0.2443
v	0.3288	0.2816
h	0.2486	0.2118
H	0.2406	0.2018

レ分布はベータ分布を多変量に拡張した分布である．目的変数 1 つ 1 つは 0 から 1 の範囲をとる値であり，合計が 1 となるのが特徴である．

本研究では Maier [23] を参考にディリクレ回帰モデルを作成した．一般化線型モデルの式を式 (1)，式 (2) に示す．

$$\log \alpha = \beta_1 e_p + \beta_2 e_a + \beta_3 e_d + \beta_4 Dur + \beta_0 \quad (1)$$

$$(y_n, y_v, y_h, y_H) \sim \mathrm{Dir}(\alpha_n, \alpha_v, \alpha_h, \alpha_H) \quad (2)$$

式 (2) 左辺の目的変数 y は無声呼気，有声呼気，無声吸気，有声吸気いずれかの構成要素の割合であり，式 (1) でモデル化する構成要素ごとの集中度パラメータ α を持つディリクレ分布に従うとする．式 (1) 右辺の β は構成要素に依存した回帰係数であり，e_p（快–不快），e_a（覚醒–睡眠），e_d（支配–服従），Dur（laughter episode の継続時間長）は説明変数である．作成したモデルの構成要素ごとの回帰係数と切片 β_0 を**表 2** に示す．表 2 から無声呼気は服従寄りの笑い声で割合が大きくなり，有声呼気は，快寄り，支配寄りの笑い声で割合が大きくなることが分かる．無声吸気は服従寄りの笑い声で割合が大きくなり，有声吸気は覚醒寄りの笑い声で割合が大きくなることが分かる．また，発話が長い場合，無声呼気，無声吸気の割合が大きくなることが分かる．

モデルの精度を求めるために Root Mean Squared Error（RMSE）を求めた．合成時に用いる際は構成要素の割合ではなく継続時間長（s）になるため，割合を継続時間長に変えて RMSE を求めた．さらに，モデルのあてはまりの良さを示すために実測値と推定値の相関係数を求めた．各笑い声の構成要素の割合を求めた際の RMSE の結果と割合を継続時間長に換算した際の RMSE の結果を**表 3** に，推定した継続時間長と正解となる継続時間長との散布図を**図 4** に示す．モデルの精度はいずれの構成要素も継続時間長で 0.2 秒以上となり，各構成要素の平均的な長さが 0.20〜0.28（標準偏差は 0.12〜0.15）であること（4.2 節を参照）

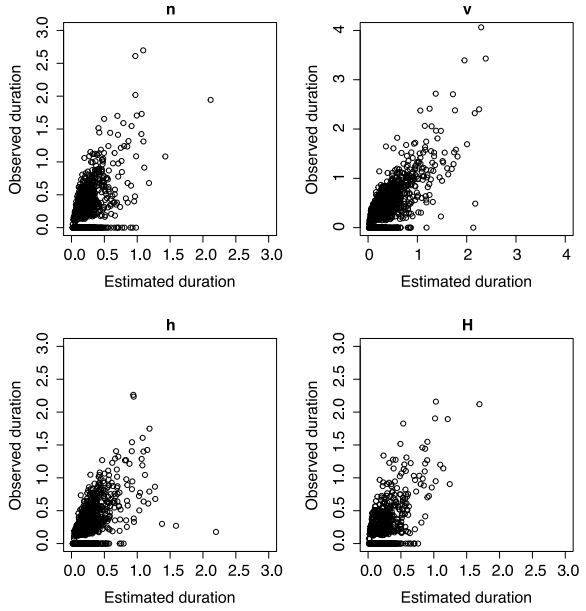

図 4 推定値と実測値の散布図（継続時間長）
Fig. 4 Scatter plot of estimated and observed duration for each laughter component.

を考慮すると，やや大きな誤差となった．一方で，実測値と推定値の相関係数は n が 0.62，v が 0.77，h が 0.65，H が 0.67 であり，ある程度の推定が可能であることを示している．図 4 に示した各構成要素の散布図を見ると，長い笑い声では誤差が大きいが，1 秒以下の典型的な笑いでは誤差は小さい．また，傾きが過小に推定されているが，これは継続時間長が 0 秒である構成要素の影響であると考えられる．したがって，1 秒未満の継続長である笑い声であるならば，このモデルである程度の推定ができるといえる．

5.2 ラベル作成

ディリクレ回帰モデルによって推定した笑い声の構成要素の割合から合成時に入力するラベルを作成する手順を**表 4** に示す．まずは，各構成要素の割合（compPer）を継続時間長（compDur）に変換する（表 4 2)-A)）．続いて，各構成要素の平均継続長と標準偏差（表 1）を参照し，各構成要素をその継続時間長が平均 $\pm 1\sigma$ の範囲となるように分割する（表 4 2)-B)-i)）．笑い声の構成要素の継続時間長と感情次元の相関を分析したところ相関がみられず（**表 5**），構成要素の継続時間長によって知覚される感情に差はないことが示されたため，推定した構成要素の継続時間長が長い場合はその平均値（meanCompDur）と標準偏差（sdCompDur）を参考に分割を行うこととした（表 4 2)-B)-i)）．また，構成要素の継続時間長がその構成要素の最小値（minCompDur）を下回る場合は合成時の要素には含めず，その構成要素の時間分は合成には含めないこととした（表 4 2)-B)-iii)）．今回は各構成要素の文脈による笑い声の自然性や感情知覚の違いを分析の対象とし

表 4 笑い声の構成要素の割合推定後のラベルの作成手順
Table 4 Algorithm for the laughter component arrangement using given component ratio.

```
1) Initialization: laughEpDur ← 9.50e7, j ← 0
2) for i = 0 to 4
  A) Estimate component's duration
    compDur ← laughEpDur ∗ compPer[i]
  B) while compDur > 0
    i) if compDur >=
       meanCompDur[i] + sdCompDur[i] + minCompDur[i]
       a) tmpDur[j] ← meanCompDur[i] +
          (rand() % (2 ∗ sdCompDur[i] + 1) − sdCompDur[i])
       b) compDur ← compDur − tempDur[j] ;
       c) Increment j
    ii) elseif compDur >= minCompDur[i]
       a) tmpDur[j] ← compDur
       b) compDur ← 0
       c) Increment j
    iii) else
       a) compDur ← 0
    iv) endif
  C) end
3) end
4) Order randomization: tmpDur[]
```

表 5 笑い声の構成要素の継続時間長と感情次元の相関係数
Table 5 Correlation coefficients between emotion dimension and duration of laughter components.

Components	Pleasantness	Arousal	Dominance
n	−0.198	−0.192	−0.181
v	−0.115	−0.123	−0.054
h	0.021	−0.009	−0.004
H	−0.021	−0.005	−0.024

ないため，各構成要素の配置順序はランダムに決定した（表 4 4））．

6. 評価実験

6.1 笑い声合成モデルの作成

データセットには OGVC に含まれている男性話者 04_MSY と女性話者 06_FWA の 2 名の話者の笑い声を用いた．04_MSY の笑い声数は 256 個であり，06_FWA の笑い声数は 237 個であった．評価実験では感情 3 次元を組み合わせた笑い声を合成するため，1 つの条件で 125 個の笑い声を用意する必要がある．そのため，他 3 条件でも同数の合成音声が必要となり，1 話者あたりが評価対象とする笑い声数は 500 個（125 個 × 4 条件）となる．評価対象とする話者が多いと音声の数も増大し被験者の負担も増えることから，今回は男性 1 名および女性 1 名の話者に限定して評価を行った．04_MSY と 06_FWA は先行研究 [7] で安定した笑い声合成が実現できることが分かっているため，今回の実験における評価対象話者とした．

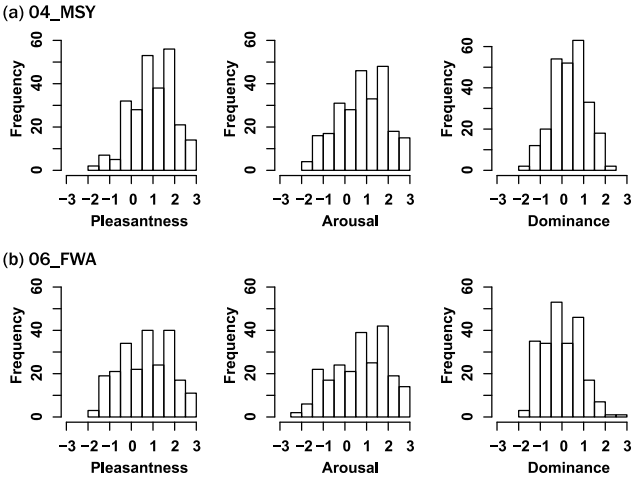

図 5 話者ごとの笑い声に対する感情次元値の分布
Fig. 5 Histograms of the emotion dimension score for each speaker's laughter in the dataset.

笑い声の感情次元値の分布を話者ごとに図 **5** に示す．詳細は 6.2 節に記述するが，この分布を持つ笑い声から実験に使用するラベルファイルを抽出する．合成器には WaveNet を利用し，笑い声の学習は話者ごとに行った．つまり，04_MSY と 06_FWA の 2 つの異なる笑い声モデルを作成している．WaveNet の学習の際には笑い声の音声ファイルと 3.2 節で説明した人手で付与した構成要素のラベルファイルを入力として用いている．なお，各手法において bout 中の構成要素の位置情報や構成要素内のサンプル点の位置情報は WaveNet に入力していない．音声ファイルは 48 kHz から 16 kHz へダウンサンプリングしている．

6.2 実験条件

本研究では，以下の 2 点について検証を行う．

(1) **Call レベルの構成要素による笑い声合成**
笑い声の構成要素を無声呼気・有声呼気・無声吸気・有声吸気の 4 種類にした方が，従来手法である bout・無声吸気・有声吸気の 3 種類の構成要素を利用するより，より自然で表現豊かな笑い声を合成可能である．

(2) **感情次元情報から推定した構成要素による笑い声合成**
笑い声の構成要素を 5 章で示した提案手法により自動で生成しても，自然性と感情の表現性を損なうことはない．

以上，2 点を検証するために，以下の 3 つの実験条件を用意した．

従来手法（baseline）：1 つ目の条件は従来手法（baseline）である Mori ら [7] の手法であり，感情次元値から笑い声の構成要素を推定したラベルを用いるのではなく，すでにコーパスに存在している構成要素のラベルを用いる．合成を行う際に用いるラベルは bout，無声吸気，有声吸気の 3 つの構成要素で構成されている．従来手法で笑い声を合成する際に使用する笑い声ラベルは，学習時に使用してい

ない笑い声の構成要素の組合せである必要がある．そのため，合成時に使用した笑い声ラベルは 2 名の話者のデータのうち学習時に使用していない話者のデータより抽出して利用した．つまり，04_MSY のモデルならば 06_FWA の笑い声ラベルを，06_FWA のモデルならば 04_MSY の笑い声ラベルを合成時に利用した．従来手法で合成する笑い声は，提案手法 2 と同数とするため，話者ごとに 125 個，合計 250 個の笑い声を用意した．また，提案手法 2 で合成する笑い声の継続時間長は laughter episode の平均値を基準としていることから，従来手法で合成する笑い声についても継続時間長は 0.95 秒を基準として笑い声ラベルを選定した（平均継続長 0.98 秒，標準偏差 0.33 秒）．

提案手法 1（proposed1）：2 つ目の条件である提案手法 1（proposed1）は bout を call レベルに分割した構成要素のラベルを用いて，従来手法よりも情報量の多い笑い声合成を実現する．ここでも，感情次元値から笑い声の構成要素を推定したラベルを用いるのではなく，すでにコーパスに存在している構成要素のラベルを用いる．合成を行う際に用いるラベルを無声呼気，有声呼気，無声吸気，有声吸気の 4 つの構成要素で構成した．提案手法 1 でも従来手法と同様に，合成時に使用した笑い声ラベルは 2 名の話者のデータのうち学習時に使用していない話者のデータより抽出して利用している．提案手法 1 で合成する笑い声も，提案手法 2 と同数とするために話者ごとに 125 個，合計 250 個の笑い声を用意した．また，従来手法で合成する笑い声についても継続時間長が提案手法 2 とほぼ同じになるように笑い声ラベルを選定した（平均継続長 0.98 秒，標準偏差 0.33 秒）．

提案手法 2（proposed2）：3 つ目の条件である提案手法 2（proposed2）では，4 章で示した手法を用いて，感情次元情報から call レベルの構成要素を推定し，笑い声を合成することを実現する．合成に用いるラベルを 4 章で示した手法により感情次元情報から笑い声の構成要素を推定することによって作成している．提案手法 2 で入力情報として扱う各感情次元は −3 から 3 の 7 段階のうち，−3，−1，0，1，3 の 5 段階とし，3 つの感情次元における 5 段階のすべての組合せである 125 通りの感情を表現した笑い声を話者ごとに合成した．笑い声の継続時間長は laughter episode の長さのヒストグラムを視察し，laughter episode が平均値付近に分布していたことを確認したうえで，平均値である 0.95 秒とした．

このほかに，各条件で合成した笑い声の評価に対する参照として，原音声（Natural）であるコーパスに含まれる人間の笑い声も聴取実験に含めた．原音声の笑い声も，提案手法 2 と同等の数とするため，話者ごとに 125 個，合計 250 個の笑い声を用意した．用意した原音声の平均継続時間長は 0.99 秒，標準偏差 0.33 秒であった．従来手法と提案手法 1 を比較することで，前述の (1) を検証する．さらに，提案手法 1 と提案手法 2 を比較することで前述の (2) を検証する．

6.3 実験環境

被験者は日本人大学生 7 名（男性 6 名，女性 1 名）である．実験は防音室を利用し，ノートパソコン（macOS）にヘッドフォン（AKG K271 MK II）をつないで行った．各被験者に笑い声 100 個を 1 セッションとしてランダムに 1 個ずつ提示し，10 セッション行った．また，各セッション終了時に 5 から 10 分ほどの休憩を挟んだ．

6.4 評価項目

評価項目は**表 6** に示す自然性，快–不快，覚醒–睡眠，支配–服従とした．自然性評価は，ITU-T Recommendation P.800 [24] を参考に，悪い（1），少し悪い（2），まあまあ良い（3），良い（4），非常に良い（5）の 5 段階とした．また，感情次元評価は原音声に付与されている評価値と同じ感情 3 次元とした．快–不快は，非常に不快（−3），かなり不快（−2），やや不快（−1），どちらでもない（0），やや快（1），かなり快（2），非常に快（3），覚醒–睡眠は，非常に睡眠（−3），かなり睡眠（−2），やや睡眠（−1），どちらでもない（0），やや覚醒（1），かなり覚醒（2），非常に覚醒（3），支配–服従は，非常に服従（−3），かなり服従（−2），やや服従（−1），どちらでもない（0），やや支配（1），かなり支配（2），非常に支配（3）の 7 段階とした．

6.5 分析方法

評価させた笑い声ごとに評価値の被験者間平均を求めてその笑い声の評価値とした．自然性評価では，話者ごとに実験条件を要因とした一元配置分散分析と Tukey-Kramer の HSD 検定を行った．感情次元評価では，合成時に用いたラベルに付与されている感情次元値[*1]と評価値との関係を見るために，各実験条件ごとにピアソンの積率相関分析を行った．また，手法間での相関係数を比較するため，合成時に用いたラベルに付与されている感情次元値および評価値からなるデータ $\mathcal{D}$ が 2 変量正規分布からの標本であ

表 6 評価項目
Table 6 Evaluation index.

項目	説明
自然性	人間らしい笑い声であるか
快–不快	話者の気分の良し悪し
覚醒–睡眠	話者の心理状態の活発さ
支配–服従	話者が会話をリードしているか

[*1] 6.2 節で述べたように，合成時のラベルおよび感情次元値は別話者のものであるが，別途実施したディリクレ回帰モデルの分析によれば感情次元と話者の交互作用はすべての組合せで有意ではなく（$p > 0.05$），別話者の感情次元値を評価に利用することに特段の問題はない．

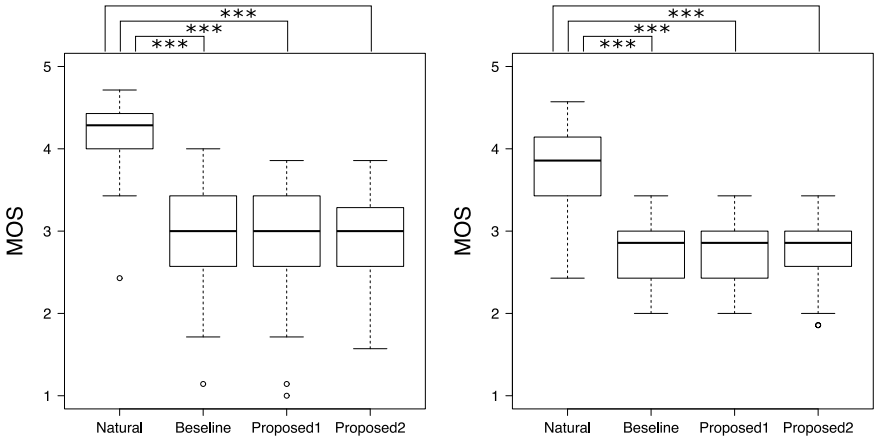

図 6 自然性評価における各評価対象の分布（左：男性話者，右：女性話者，$***p < 0.001$）
Fig. 6 Results of naturalness evaluation for each condition (left: male speaker, right: female speaker, $***p < 0.001$).

ると仮定し，従来手法の母相関係数を ρ_0，提案手法 1 および 2 の母相関係数を ρ_1，ρ_2 として，2 手法の相関係数の差 $\Delta\rho$ の事後分布 $p(\Delta\rho|\mathcal{D})$ をマルコフ連鎖モンテカルロ法（MCMC 法）により近似的に求め，そこから $\Delta\rho$ が 0 よりも大きい確率 $p(\Delta\rho > 0|\mathcal{D}) = \int_0^\infty p(\Delta\rho|\mathcal{D})d\Delta\rho$ を推定した．

7. 実験結果

自然性評価の結果を**図 6** に示す．図 6 は話者ごとに各笑い声の MOS（平均オピニオンスコア）の分布を箱ひげ図により示している．04_MSY における各条件の平均 MOS 値は，原音声が 4.18，従来手法が 2.92，提案手法 1 が 2.93，提案手法 2 が 2.95 である．また，06_FWA における各条件の平均 MOS 値は，原音声が 3.79，従来手法が 2.75，提案手法 1 が 2.76，提案手法 2 が 2.81 である．一元配置分散分析を行った結果，04_MSY が $F(3, 496) = 231.00$，$p < 0.001$，06_FWA が $F(3, 496) = 291.27$，$p < 0.001$ で実験条件の主効果が有意であった．多重比較を行った結果，04_MSY と 06_FWA ともに原音声が各合成手法に比べ自然性が有意に高く（$p < 0.001$），合成手法間に統計的な有意差は示されなかった．

感情次元評価から相関係数を求めた結果を**表 7** に，従来手法より提案手法 1 および 2 の相関係数が高い確率 $p(\rho_1 > \rho_0|\mathcal{D})$ および $p(\rho_2 > \rho_0|\mathcal{D})$ を推定した結果を**表 8** に示す．表 7 より，提案手法 1 は従来手法に比べ，04_MSY および 06_FWA ともにすべての感情次元で相関係数が高いことが分かる（04_MSY：快–不快が +0.13，覚醒–睡眠が +0.15，支配–服従が +0.26，06_FWA：快–不快が +0.09，覚醒–睡眠が +0.14，支配–服従が +0.18）．表 8 より，6.5 節で述べた仮定の下では，「提案手法 1 は従来手法に比べ相関係数が高い」という仮説はそれぞれ 77%から 99%の確率で正しいと解釈できる．また，表 7 より，提案手法 2 は従来手法に比べ，04_MSY では支配–服従次元で，06_FWA では覚醒–睡眠次元および支配–服従次元で相関係数が高いことが分かる（04_MSY：支配–服従が +0.09，06_FWA：覚醒–睡眠が +0.17，支配–服従が +0.06）．表 8 より，これらの話者および感情次元では「提案手法 2 は従来手法に比べ相関係数が高い」という仮説は 83%から 91%の確率で正しいと解釈できる．また，提案手法 1 に比べ，提案手法 2 は 06_FWA では覚醒–睡眠次元で相関係数が高い（06_FWA：覚醒–睡眠が +0.03）ものの，6.5 節で述べた方法で手法間

表 7 話者ごとの各条件における相関係数の結果（$\dagger p < 0.1$，$*p < 0.05$，$**p < 0.01$，$***p < 0.001$）
Table 7 Correlation coefficients of input emotion dimension score and evaluated emotion dimension score in each condition ($\dagger p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

Pleasantness	04_MSY	06_FWA
baseline	0.26**	0.13
proposed1	0.39***	0.22*
proposed2	0.03	0.15
Arousal	04_MSY	06_FWA
baseline	0.27**	0.03
proposed1	0.42***	0.17 †
proposed2	0.11	0.20*
Dominance	04_MSY	06_FWA
baseline	0.15**	−0.03
proposed1	0.41***	0.15
proposed2	0.24**	0.03

表 8 従来手法より提案手法 1 および 2 の相関係数が高い確率
Table 8 Probability that the correlation coefficient of proposal 1 or 2 is greater than that of the baseline.

	$p(\rho_1 > \rho_0\|\mathcal{D})$		$p(\rho_2 > \rho_0\|\mathcal{D})$	
Emotion	04_MSY	06_FWA	04_MSY	06_FWA
Pleasantness	0.87	0.77	0.04	0.56
Arousal	0.90	0.87	0.10	0.91
Dominance	0.99	0.94	0.83	0.83

の相関係数を比較したところ，「提案手法 2 が提案手法 1 より相関係数が高い」という仮説が正しいと解釈できる確率 $p(\rho_2 > \rho_1|\mathcal{D})$ は 57%であった．

8. 考察

8.1 Call レベルの構成要素による笑い声合成

提案手法 1 では自然性評価で従来手法と統計的な有意差がなく，感情次元評価で従来手法に比べ高い相関がみられた．従来手法と提案手法 1 の違いは構成要素の種類である．従来手法が呼気による音響イベントである call を複数個含む bout を呼気の入力の単位としているのに対し，提案手法 1 ではその bout 内の複数の call を有声と無声で区別し，より細かな単位で柔軟な入力を可能としている．そのため，call レベルの情報を用いることで従来手法と同程度の自然性を保持しつつ，感情をより正確に知覚させることのできる笑い声を合成可能であることが明らかとなった．これは，逆に，bout に含まれている無声呼気と有声呼気の感情次元に対する知覚特性が異なっているために，無声呼気および有声呼気を柔軟に指定することのできる call レベルの入力の方がより正確に感情を知覚させる笑い声を合成することができたからであると考えられる．bout 単位で笑い声を生成すると，bout に含まれる構成要素の有声/無声を制御することができず，感情知覚の精度を落としてしまう可能性が示唆される．

8.2 感情次元情報から推定した構成要素による笑い声合成

提案手法 2 では自然性評価で従来手法および提案手法 1 と統計的に有意な差は示されなかった．一方で，感情次元評価では，04_MSY の支配次元と 06_FWA の覚醒次元で従来手法よりも強い相関を示すとともに 06_FWA の覚醒次元で提案手法 1 よりもやや強い正の相関を示した．感情次元の情報を利用して笑い声の構成要素を決定しても，従来手法と同等の自然性を保持したまま，従来手法よりもいくつかの感情をより正確に知覚させることができる可能性が示唆される．

一方で，提案手法 2 では 06_FWA の覚醒次元以外の感情次元では提案手法 1 より強い相関係数を示すことはなかった．提案手法 1 と 2 の違いは，構成要素の配置順序の決定に，コーパスに存在する構成要素をそのまま使うか，コーパスを参照せずにモデルにより推定するかどうかである．提案手法 2 では構成要素の並び順を考慮していない．そのため，合成する笑い声に各構成要素がどのくらいの割合で含まれているかだけでなく，その配置順序を推定することで，感情次元知覚を改善できる可能性がある．

9. おわりに

本研究では笑い声の構成要素ごとに感情知覚特性が異なることを利用し，感情情報から笑い声の構成要素を推定するモデルを作成して，推定した構成要素を入力として自然な笑い声が合成可能であるか，およびその合成した笑い声から指定した感情が知覚可能であるか検証した．

その結果，WaveNet を用いた笑い声合成では call レベルの情報を利用することで自然性を保持しつつ，より知覚させたい感情を表現する笑い声を合成可能であることを明らかにした．また，感情情報を用いたディリクレ回帰により各構成要素の割合を推定して，その要素列を決定しても，自然性を損なうことなく，指定した感情次元の情報を持つ笑い声が合成可能であることを示した．本研究の結果は，音声学や音響学の知識がないユーザが感情情報を直感的に操作することで自然で表現豊かな笑い声を合成可能なインタフェースの実現に貢献するものである．

本研究の結果より笑い声の構成要素の割合だけでなく，各構成要素の配置順序も感情知覚に影響する可能性があることが示唆されたため，今後は構成要素の配置順序を自動で推定するモデルの構築を検討する予定である．

なお，今回は純粋な笑い声を対象とした合成手法について検討したが，2.1 節に示したとおり音声に現れる笑いには笑い声の特徴を持った発話である speech laugh や笑顔に聞き取れる smiled speech があり，この 2 つの笑いについては検討していない．自然で表現力の高い音声合成を実現するためには，speech laugh と smiled speech を合成することは不可欠である．現在，speech laugh の音声学的な発生機序について分析が進められており，人は発話中のどこで笑い始めるかが解明されはじめている [25]．まずは笑いに関する様々な音声現象を解明し，表現力の高い音声合成の実現を目指したい．

参考文献

[1] Shen, J., Pang, R., Weiss, R.J., Schuster, M., Jaitly N., Chen Z., Zhang, Y., Wang, Y., Skerrv-Ryan, R., Saurous, R.A., Agiomvrgiannakis, Y. and Wu, Y.: Natural TTS Synthesis by Conditioning WaveNet on MEL Spectrogram Predictions, *2018 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), pp.4779–4783, DOI: 10.1109/ICASSP.2018.8461368 (2018).

[2] van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A. and Kavukcuoglu, K.: WaveNet: A Generative Model for Raw Audio, arXiv:1609.03499v2 [cs.SD] (2016).

[3] Lorenzo-Trueba, J., Eje Henter, G., Takaki, S., Yamagishi, J., Morino, Y. and Ochiai, Y.: Investigating different representations for modeling and controlling multiple emotions in DNN-based speech synthesis, *Speech Communication*, Vol.99, pp.135–143, DOI: 10.1016/j.specom.2018.03.002 (2018).

[4] Nose, T. and Kobayashi, T.: An intuitive style control technique in HMM-based expressive speech synthesis using subjective style intensity and multiple-regression global variance model, *Speech Communication*, Vol.55, No.2, pp.347–357, DOI: 10.1016/j.specom.2012.09.003 (2013).

[5] Devillers, L. and Vidrascu, L.: Positive and negative emotional states behind laughs in spontaneous spoken dialogs, *Interdisciplinary Workshop on The Phonetics of Laughter Saarbrücken*, pp.37–40 (2007).
[6] Tanaka, H. and Campbell, N.: Classification of social laughter in natural conversational speech, *Comput. Speech Lang.*, Vol.28, pp.314–325 (2014).
[7] Mori, H. Nagata, T. and Arimoto, Y.: Conversational and Social Laughter Synthesis with WaveNet, *Proc. Interspeech 2019*, pp.520–523, DOI: 10.21437/Interspeech.2019-2131 (2019).
[8] Tits, N., El Haddad, K. and Dutoit, T.: Laughter Synthesis: Combining Seq2seq Modeling with Transfer Learning, *Proc. Interspeech 2020*, pp.3401–3405, DOI: 10.21437/Interspeech.2020-1423 (2020).
[9] Bachorowski, J.A. and Owren, M.J.: Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect, *Psychological Science*, Vol.12, No.3, pp.252–257, DOI: 10.1111/1467-9280.00346 (2001).
[10] 小山俊樹，有本泰子：ゲームプレイ中に表出した社会的な笑い声のラベリングとその音響分析，日本音響学会 2020 年春季研究発表会講演論文集，pp.827–828 (2020).
[11] Szameitat, D.P., Alter, K., Szameitat, A.J., Darwin, C.J., Wildgruber, D., Dietrich, S. and Sterr, A.: Differentiation of emotions in laughter at the behavioral level, *Emotion*, Vol.9, No.3, pp.397–405, DOI: 10.1037/a0015692 (2009).
[12] 今西利於，小山俊樹，有本泰子：笑い声の感情次元知覚に対する呼気音，吸気音の影響，日本音響学会 2020 年秋季研究発表会講演論文集，pp.787–788 (2020).
[13] Urbain, J., Çakmak, H., Charlier, A., Denti, M., Dutoit, T. and Dupont, S.: Arousal-driven synthesis of laughter, *IEEE Journal on Selected Topics in Signal Processing*, Vol.8, No.2, pp.273–284, DOI: 10.1109/JSTSP.2014.2309435 (2014).
[14] Russell, J.A.: A Circumplex Model of Affect, *Journal of Personality and Social Psychology*, Vol.39, No.6, pp.1161–1178, DOI: 10.1037/h0077714 (1980).
[15] Mehrabian, A. and O'Reilly, E.: Analysis of personality measures in terms of basic dimensions of temperament, *Journal of Personality and Social Psychology*, Vol.38, No.3, pp.492–503 (1980).
[16] Truong, K.P., Trouvain, J. and Jansen, M.: Towards an annotation scheme for complex laughter in speech corpora, *Proc. Interspeech 2019*, pp.529–533 (2019).
[17] 森　大毅，有本泰子，永田智洋：複数の会話コーパスを対象とした笑い声イベントのアノテーション，日本音響学会 2017 年秋季研究発表会講演論文集，pp.217–218 (2017).
[18] Nagata, T. and Mori, H.: Defining Laughter Context for Laughter Synthesis with Spontaneous Speech Corpus, *IEEE Trans. Affective Computing*, Vol.11, No.3, pp.553–559, DOI: 10.1109/TAFFC.2018.2813881 (2020).
[19] Kumar, K., Kumar, R., de Boissiere, T., Gestin, L., Teoh, W.Z., Sotelo, J., de Brebisson, A., Bengio, Y. and Courville, A.: MelGAN: Generative Adversarial Networks for Conditional Wavefrom Synthesis, arXiv:1910.06711v3 [eess.AS] (2019).
[20] Szaameitat, D.P., Drawin, C.J., Wildgruber, D., Alter, K. and Szameitat, A.J.: Acoustic correlates of emotional dimensions in laughter: Arousal, dominance, and valence, *Cognition & Emotion*, Vol.25, No.4, DOI: 10.1080/02699931.2010.508624 (2011).
[21] Arimoto, Y., Kawatsu, H., Ohno, S. and Iida, H.: Naturalistic emotional speech collection paradigm with online game and its psychological and acoustical assessment, *Acoustical Science and Technology*, Vol.33, No.6, pp.359–369, DOI: 10.1250/ast.33.359 (2012).
[22] Boersma, P. and Weenink, D.J.M.: PRAAT, A system for doing phonetics by computer, *Glot International*, Vol.5, No.9, pp.341–347 (2001).
[23] Maier, M.J.: DirichletReg: Dirichlet Regression for Compositional Data in R, Research Report Series/Department of Statistics and Mathematics, WU Wirtschaftsuniversität Wien, Vol.125 (2014).
[24] Methods for subjective determination of transmission quality, ITU-T Recommendation P.800 (1998).
[25] 有本泰子，阿部将士，津田逸成：Speech laugh が発生しやすい調音に関する音声学的分析，日本音響学会 2021 年秋季研究発表会講演論文集，pp.817–818 (2021).

有本 泰子

1973 年生．1996 年青山学院大学文学部英米文学科卒業．2004 年東京工科大学メディア学部メディア学科卒業．2007 年同大学大学院博士前期課程修了．2010 年同大学院博士後期課程単位取得退学．博士（工学）．2010 年科学技術振興機構戦略的創造研究推進事業（JST，ERATO）岡ノ谷情動情報プロジェクト研究員．2010 年理化学研究所脳科学総合研究センター客員研究員．2015 年芝浦工業大学工学部共通学群情報科目特任准教授．2017 年帝京大学理工学部情報電子工学科講師．2019 年千葉工業大学情報科学部情報工学科准教授．対話音声を中心とした感情認識・感情知覚・音声合成の研究に従事．日本音響学会，日本音声学会，ISCA 各会員．

今西 利於

1998 年生．2021 年千葉工業大学情報科学部情報工学科卒業．同年株式会社エーアイ入社．在学中は笑い声合成の研究に従事．日本音響学会会員．

森 大毅　（正会員）

1971 年生．1993 年東北大学工学部通信工学科卒業．1998 年同大学大学院博士後期課程修了．博士（工学）．1998 年東北大学助手．1999 年宇都宮大学助手．2005 年同大学助教授．現在，同大学大学院工学研究科准教授．音声合成，音声対話，医療福祉工学の研究に従事．電子情報通信学会，日本音響学会，人工知能学会，ISCA 各会員．