Vol.2019-MUS-123 No.39
Vol.2019-SLP-127 No.39
2019/6/22

深層ニューラルネットワークを用いた波形接続型感情音声合成のための感情制御法

大谷　大和[1,a)]　松永　悟之[1]　平井　啓之[1]

概要：本稿では深層学習を用いた波形接続型感情音声合成のための感情制御法について述べる．従来の波形接続型感情音声合成では，1) 素片単位での混合が困難であるため，中間的な感情表現が乏しい，2) 入力された感情強度に従い素片の感情の種類を切り替えるため，感情による声質の変化が不連続になるといった問題があった．これらの問題を解決するために，提案手法では深層ニューラルネットワーク（DNN）を用いて，平静音声のスペクトル特徴量と感情強度から感情音声と平静音声の差分スペクトルを予測し，これを平静の素片に畳み込むことで所望の感情強度の感情素片を生成する．また，入力感情強度に応した差分スペクトル特徴量を予測可能にするため，データ拡張により感情強度に対応した差分スペクトル特徴量を生成し，これらを学習に用いることで所望の制御則を DNN に埋め込む．実験的評価では，従来手法と比較して滑らかな感情制御ができていることを確認した．

キーワード：テキスト音声合成，波形接続型感情音声合成，感情制御，深層学習，テータ拡張

Emotion manipulation for unit-selection-based speech synthesis using deep neural network

Abstract: This paper describes a novel emotion manipulation method for unit-selection-based speech synthesis (USS) using a deep neural network. Our conventional unit-selection-based emotional speech synthesis (USES) includes two weaknesses; 1) it is poor at mixed emotional expressions because it is difficult to generate interpolated units, and 2) variations of emotional voice quality are discontinuous because emotional unit set are changed based on input emotion intensities. To solve these problems, the proposed method predicts spectral differentials between emotional and neutral speech from input emotional intensities and neutral spectral features using the deep neural network (DNN). Then the emotional units are generated by convolution of neutral ones with predicted spectral differentials. Moreover, in order to generate spectral differentials corresponding with input emotional intensities, we introduce data augmentation technique to training of DNNs. Experimental results show that the proposed method achieves smooth manipulations of emotional intensities compared with the conventional USES.

Keywords: Text-to-speech, unit-selection-based emotional speech synthesis, emotion manipulation, deep learning, data augmentation.

1. はじめに

テキスト音声合成 [1] はある入力された文字列をその内容を表した音声信号に変換する技術である．近年，音声合成はその技術的な発展により，肉声感の向上に関する研究のほか，表現の多様化についても広く研究が進められている．そのような音声合成の多様化の技術のひとつとして感情音声合成 [2] があり，動画などのコンテンツ作成やロボット対話システムなどで利用されている．

多くの感情音声合成の研究では，通常発声（平静）や，喜び，怒り，悲しみなどを表現した感情音声からなるコーパスを用意し，これらを用いて音声合成システムを構築している．近年では，隠れマルコフモデル（HMM）や深層ニューラルネットワーク（DNN）を用いたパラメトリック音声合成 [3], [4], [5] に基づいた感情音声合成システムの研究が行われている [6], [7], [8], [9]．これらの手法では各感

[1] 株式会社エーアイ
[a)] ohtani@ai-j.jp

ⓒ 2019 Information Processing Society of Japan

Vol.2019-MUS-123 No.39
Vol.2019-SLP-127 No.39
2019/6/22

情音声を合成できるほか，各感情モデルのモデルパラメータや出力音響特徴量を線形結合することで，複数の感情を混合した合成音声を生成することが可能となる．また，感情音声の収録のない場合でも，既存の感情音声から感情に関するパラメータを学習，予測し，これを任意の話者に適用することで，感情のない話者にも感情合成音声を生成する研究が進められている [10], [11].

我々は，これまでに波形接続型音声合成に基づいた感情音声合成システムを開発し，コンシューマ向けおよび法人向け製品・サービスを展開している．本手法ではパラメトリック音声合成と同様に，平静を含めた各感情音声を用いる．各感情音声からそれぞれの韻律辞書と素片辞書を構築し，これらを用いて感情音声合成を実現している．本手法では収録音声波形をそのまま用いて合成するため，肉声感の高い感情音声を実現できる．しかしながら，パラメトリック音声合成のように音声波形をパラメータ化していないため，複数の感情音声波形を混合するのが非常に困難であり，パラメトリック音声合成と比較して感情の表現能力が乏しくなる．また，入力された感情強度に従って，合成に用いる素片辞書の感情を切り替えるため，滑らかな感情制御ができない問題がある．

本稿では，これらの問題を解決するために，ディープニューラルネットワーク (DNN) に基づく感情音声制御法を提案する．提案法では，声質変換技術 [12], [13], [14], [15] のひとつである差分スペクトル補正 [16] を応用し，DNN を用いて，平静音声のスペクトル特徴量と感情強度から平静音声と感情音声の差分スペクトル特徴量を予測し，これを平静の音声素片に畳み込むことで所望の強さの感情の合成音声を得る．また，ユーザが指定した感情強度に対応した差分スペクトル特徴量を生成できるようにするために，データ拡張 [17] を導入する．データ拡張とは，学習データに偏りや欠損がある場合に，あらかじめ何らかの処理を行い学習データ数を増やす手法である．本手法では，ある制御則に従い，入力感情強度に応じた差分スペクトル特徴量を生成し，これらを学習データとして用いることで，所望の感情強度の差分スペクトル特徴量を予測する DNN の実現をめざす．

本稿の構成は次のとおりである．2 節では，波形接続型感情音声合成の概要について述べる．3 説では提案手法である DNN を用いた感情制御が可能な波形接続型音声合成について述べる．5 節で実験的評価について述べ，6 節にて本稿をまとめる．

2. 波形接続型感情音声合成の概要

我々の感情音声合成システムの構成を図 1 に示す．本システムでは，各感情音声コーパスからそれぞれの韻律辞書と素片辞書を構築する．韻律辞書は統計的手法を用いて構築されており，韻律情報（基本周波数，継続長，信号のパ

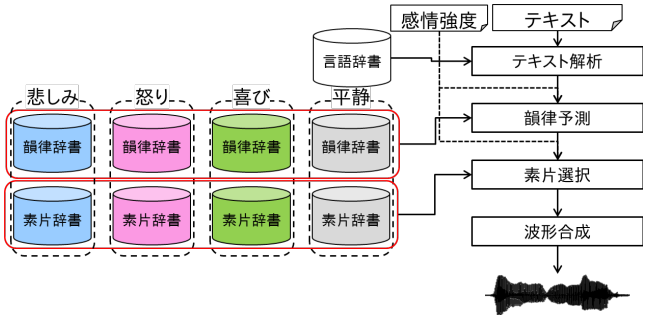


図 1　波形接続型感情音声合成

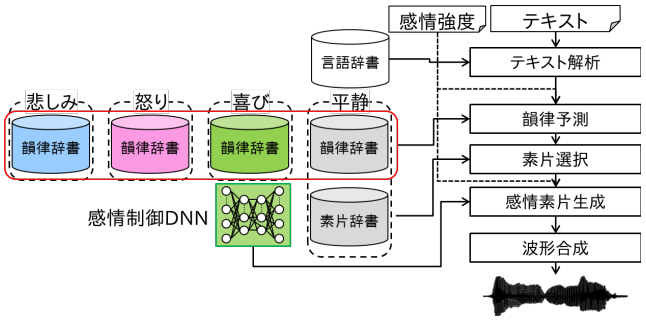


図 2　DNN を用いた波形接続型感情音声合成

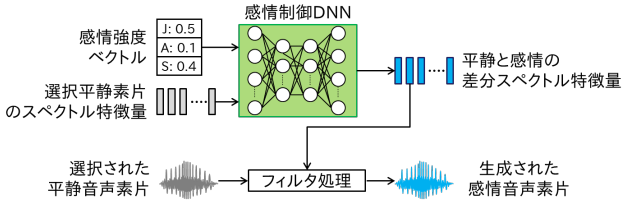


図 3　感情素片生成部の詳細

ワー）を予測する．素片辞書は収録音声に対して，音素インデックスや音素境界位置，音響特徴量など，素片選択および波形合成時に必要な情報を付与しデータベース化したものである．本感情音声合成システムでは，入力されたテキストからテキスト解析により音素や品詞などの言語特徴量（コンテキスト）を生成しする．韻律予測では，まず生成したコンテキストから各感情の韻律辞書を用いて感情ごとの韻律を予測する．ユーザが入力した感情強度に基づいて，予測された各感情韻律を線形結合し，所望の感情韻律を得る．次に素片選択部では，まず入力感情強度に従い合成に用いる感情の素片辞書を選択する．その後，波形合成部において予測感情韻律に従い，Pitch-synchronous overlap and add (PSOLA) [18] により選択素片を変形し接続するで所望の感情合成音声を得る．

3. DNN を用いた波形接続型感情音声合成

3.1 概要

提案システムの概要を図 2 を示す．本システムでは，韻律予測部まで，2 節で述べた従来の波形接続型感情音声合

ⓒ 2019 Information Processing Society of Japan

成と同様の処理を行うため，韻律辞書についてはシステムで用いるすべての感情の辞書を用意する．一方，素片辞書については平静のみを用い，素片選択部で予測韻律に対応した平静素片を選択した後，感情素片生成部において事前に学習した感情制御 DNN を用いて所望の感情素片の生成する．

感情素片生成部の概要を図 3 に示す．本処理部は，差分スペクトル補正による声質変換 [16] を応用しており，図中の感情制御 DNN が声質変換における変換モデルに相当する．感情制御 DNN $f(\cdot)$ は次式のような平静音声のスペクトル特徴量 $\boldsymbol{x}_t$ と各感情の強度を表した感情強度ベクトル $\boldsymbol{w}$ を入力とし，入力された感情強度に対応した平静音声と感情音声の差分スペクトル特徴量 $\boldsymbol{d}_t = \boldsymbol{y}_t - \boldsymbol{x}_t$ ($\boldsymbol{y}_t$: 感情音声のスペクトル特徴量) を予測することを目的とした DNN である．

$$\tilde{\boldsymbol{d}}_t = f(\boldsymbol{x}_t, \boldsymbol{w}) \tag{1}$$

本システムでは，この DNN に対して，素片選択部で選ばれた平静素片に対応するスペクトル特徴量とユーザが指定した感情強度を入力し，得られた差分スペクトル特徴量を予測する．この差分スペクトル特徴量を選択された平静素片に畳み込むことで，所望の感情強度をもつ感情音声の素片を生成する．

生成後，従来システムと同様に PSOLA により波形を変形，接続し，所望の感情強度の合成音声波形を得る．

次節において，本システムの主要となるデータ拡張を用いた感情制御 DNN の学習について述べる．

3.2 データ拡張による感情制御 DNN の学習

感情制御 DNN の学習では，平静音声のスペクトル特徴量と感情音声のスペクトル特徴量からなるパラレルデータを用いる．パラレルデータを用いる場合，入力話者と出力話者で同一内容の発話データが必要となるが，本システムで用いる感情音声のコーパスと平静音声のコーパスの間において，同一内容発話が存在しない．そのため，本稿では従来システムを用いて感情音声と同一の継続長と発話内容をもつ平静の合成音声を用意し，これらを用いてパラレルデータを構築している．また，本稿では，事前に平静音声と感情音声の差分を求め，平静音声のスペクトル特徴量および感情強度から差分スペクトル特徴量を直接予測するように学習する．

通常，感情コーパスは感情の強さが一定であることが前提で構築されているため，学習時の感情強度はいずれかの感情強度が 1.0 で残りが 0.0 となる 1-hot ベクトルような形式で学習を行うことが考えられる．しかしながら，このように学習すると，すべての感情強度を 0.0 とした場合やいずれかの感情強度を 0.3 や 0.7 といった中間的な値を指定した場合，学習時にはこれらの場合に対応したデータが使われていないため，感情制御 DNN は指定した感情強度に対応した差分スペクトル特徴量を生成することができず，その結果，合成音声の感情制御が適切にできない．そこで本稿ではデータ拡張 [17] によりこの問題を解決する．この手法では入力する感情強度に対応する差分スペクトル特徴量を人工的に生成し，これらを学習に用いる．これにより DNN に所望の感情制御則を埋め込むことができ，DNN が感情強度に対応した差スペクトル特徴量を予測できるようにする．本システムでは，感情制御則として入力された感情強度に対して線形に差分スペクトル特徴量が変化するものを用いる．これを実現するために，次式を用いてデータを生成し，データ拡張を行う．

$$\hat{\boldsymbol{d}} = \sum_{e=1}^{E} w_e \boldsymbol{d}^{(e)} \tag{2}$$

ここで w_e は e 番目の感情に対する感情強度，$\boldsymbol{d}^{(e)}$ は e 番目の差分スペクトル特徴量，および $\hat{\boldsymbol{d}}$ は人工的に作成した差分スペクトル特徴量である．本稿では，1 バッチごとに式 (2) を適用してデータ拡張を行う．また，データ拡張で用いる感情強度ベクトルは以下のものを用いる．

- 1-hot ベクトル
- すべて 0.0 としたベクトル
- 0.0〜1.0 で乱数で生成したベクトルを複数個

このように感情制御 DNN を学習することで，所望の感情制御性能をモデルに埋め込むことができる．

4. 実験的評価

4.1 実験条件

本評価では，エーアイにおいて独自に収録した 4 名の女性話者の平静，喜び，怒り，および悲しみのコーパスを用い，喜び，怒り，悲しみのそれぞれ感情コーパスから学習データ 200 文，検定データを 50 文，評価データを 50 文，それぞれ重複しないように選択した．実験では，44.1 kHz サンプリング，16 bit 量子化の音声データを用いている．スペクトル特徴量として，WORLD [19] を用いて，FFT 点数 2048，フレームシフトを 5 ms に設定しパワースペクトルを抽出し，SPTK [20] により伸縮係数 0.59 として 0 24 次のメルケプストラムを求めた．なお，メルケプストラムの次数は予備実験により決定した．また，選択素片に対するフィルタ処理には MLSA (Mel-log spectrum approximation) フィルタ [21] を用い，予測された差分メルケプストラムを畳み込んでいる．

本稿で用いる感情制御 DNN は，入力感情強度と平静のメルケプストラムの静的・動的特徴量を結合したものを入力とし，平静音声と感情音声の差分メルケプストラムを出力する．ネットワーク構造はフィードフォワードネットワークを用い，予備実験により隠れ層の数を 4, ノード数を 1024 とした．DNN の学習フレームワークとして Keras [22] を

Vol.2019-MUS-123 No.39
Vol.2019-SLP-127 No.39
2019/6/22

用い，Tensorflow [23] をバックエンドとした．DNN の最適化では Adam [24] を用いた．Adam のパラメータは文献内で提案されている推奨値を用い，1 発話ごとに DNN のパラメータを更新した．実験では平均二乗誤差 (MSE) 基準で学習したモデルと，敵対的学習 (GAN) [25] により学習したモデルを用意した．なお，敵対的学習では文献 [26] で提案されている生成誤差と識別誤差両方を考慮したものを用い，また GAN の種別として，LSGAN [27] を用いた．本稿では LSGAN で利用する識別器として，隠れ層 3，ノード数 1024 のフィードフォワードネットワークを採用した．

4.2 データ拡張による制御性付与の効果

提案法のデータ拡張の効果を確認するために客観評価を行った．評価では，3.2 節のデータ拡張に関して，学習に用いる感情強度をいくつかの条件で学習した感情制御 DNN を用意し，予測差分スペクトル特徴量が入力感情強度とどの程度一致しているかを調べた．感情制御 DNN の学習における感情強度の条件は以下のとおりである．

(a) 1-hot ベクトルのみ
(b) (a)+ゼロベクトル
(c) (b)+乱数生成した感情強度ベクトル 1 種類
(d) (b)+乱数生成した感情強度ベクトル 5 種類
(e) (b)+乱数生成した感情強度ベクトル 10 種類

なお，感情強度の条件 (a) はデータ拡張をしない場合と同義であり，また，本稿で用いる感情コーパスはすべて異なる発話であるため，学習で用いた感情強度ベクトルは少なくとも 2 つの要素はゼロとなっている．本実験の評価尺度として，入力感情強度 w と次式で定義する予測差分スペクトル特徴量から計算されるスペクトル感情強度 r の二乗平均平方根誤差（RMSE）を用いた．

$$r = \sum_{k=1}^{K} \left| \frac{\tilde{d}_{w}^{(k)}}{\tilde{d}_{w=1}^{(k)}} \right| \qquad (3)$$

ここで K は差分スペクトル特徴量の次元の数であり，$\tilde{d}_{w=1}^{(k)}$ は $w = 1$ としたときの予測差分スペクトル特徴量である．各図において点線は理想的に学習した場合の軌跡であり，実線は感情制御 DNN の予測値から計算された相対強度である．また，評価では入力感情強度を感情ごとに 0.0 から 1.0 の間で 0.05 刻みで予測差分スペクトル特徴量を生成し RMSE を求めている．

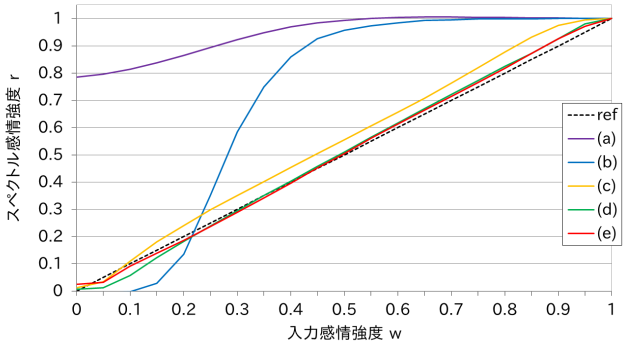
表 1 に学習基準ごとの入力感情強度と相対感情強度の RMSE を示す．データ拡張をしない場合（条件 (a)）では，いずれの学習基準の場合においても RMSE が大きい．一方，データ拡張を行うことで，RMSE が低減されていくことがわかる．また GAN のほうが MSE よりも RMSE がより低減されているのが確認できる．

次に各入力感情強度に対応するスペクトル感情強度の関係性を図 4 に示す．凡例の “ref” は本システムにおいて目標とする入力感情強度とスペクトル感情強度の関係性を図示したものであり，残りは各データ拡張の条件での関係性を図示したものである．条件 (a) の場合は，MSE，GAN のいずれの場合においてもスペクトル感情強度が “ref” から大きく外れており，感情制御がほとんどできないことがわかる．また，条件 (b) の場合は，$w = 0.0$ のときはスペクトル感情強度も 0.0 付近の値となっている一方，0.0 から 1.0 の間の変化が S 字曲線を描いており，感情制御はできるものの，システムの意図通りの制御が難しいことがわかる．乱数生成した感情強度ベクトルを加えた場合（(c)〜(e)）では，入力感情強度とスペクトル感情強度の関係性が “ref” に近づいていることから，本手法によるデータ拡張の効果があることが確認できる．さらに GAN ではほぼ理想通りの関係性となっており，適切な感情制御が実現できていることがわかる．

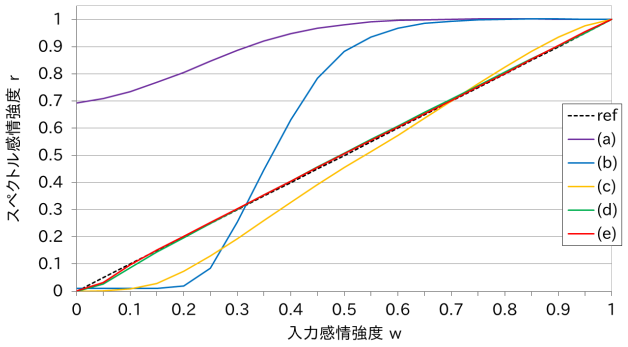
以上より，提案するデータ拡張法により，適切に感情制御が可能な DNN が得られることが確認できた．

表 1 入力感情強度とスペクトル感情強度の RMSE

データ拡張の条件	学習基準 MSE	学習基準 GAN
(a)	0.517	0.480
(b)	0.283	0.227
(c)	0.0542	0.0679
(d)	0.0215	0.00819
(e)	0.0150	0.00536



(a) 学習基準: MSE



(b) 学習基準: GAN

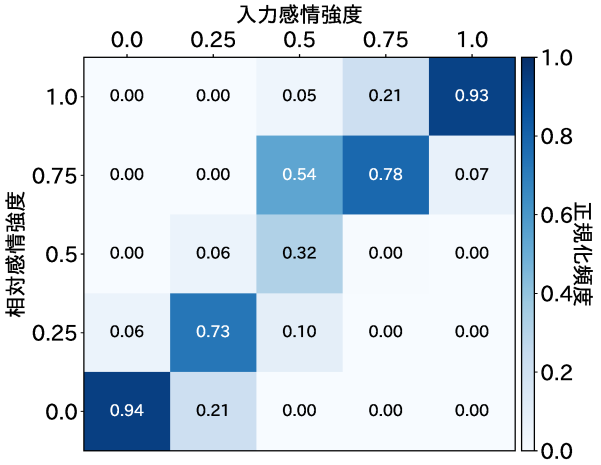
図 4 入力感情強度とスペクトル感情強度の関係

ⓒ 2019 Information Processing Society of Japan

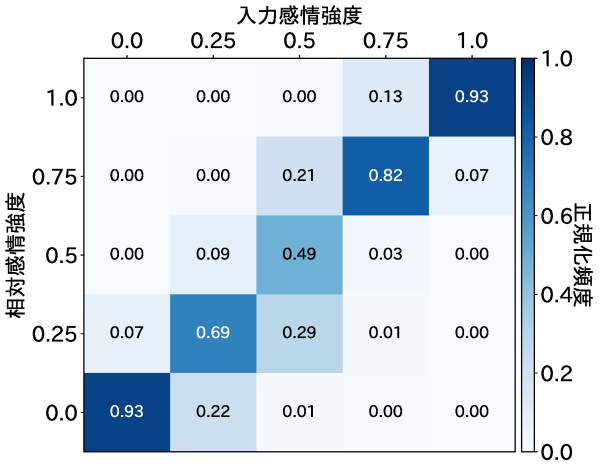
4.3 主観評価

次に従来の波形接続型感情音声合成と本稿で提案する感情音声合成の感情制御性を評価するために主観評価実験を実施した．本実験で用いる感情制御 DNN はデータ拡張の条件を (e) とし，GAN により学習したものを用いた．評価では感情毎に入力感情強度を 0.0，0.25，0.5， 0.75，1.00 とし，評価データから 3 文を選び，現行製品および本開発システムを用いて感情合成音声を生成した．評価方法では，まず評価者に各システム毎に平静の合成音声，感情強度 1.0 の感情合成音声を参照音声として聴いてもらった後，上記のうちひとつの感情強度の評価用合成音声を評価者に聴いてもらった．その後，評価音声がふたつの参照音声のうちどちらに近いかという相対的な感情強度を 0.0 から 1.0 で 0.25 刻みの 5 段階評価 (0.0: 平静音声, 1.0:感情音声) で評価者に評価してもらった．評価者数は 6 名で各評価者で 360 サンプル評価した．

Figure 5 に実験結果のヒートマップを示す．図中のセルはある入力感情強度で合成した感情合成音声の相対感情強度の正規化頻度を表しており，色が濃いものほどよく選ばれていることを示している．従来の波形接続型感情音声合成では入力感情強度が 0.5 のとき，相対感情強度が 0.75 のところが最も選択されている．これは，従来のシステムでは感情強度が 0.5 のときに素片辞書が平静音声から感情音声への切り替えが生じ，急激に声質が変化していることが原因であることがわかる．一方，提案システムでは入力感情強度が 0.5 のときは相対感情強度も 0.5 が最も選択されており，感情の制御が適切に行われていることがわかる．以上より，提案システムは従来の波形接続型感情音声合成と比べ，滑らかに制御できていると考えられる．



(a) 従来の波形接続型感情音声合成システム



(b) 提案する波形接続型感情音声合成システム

図 5 感情の制御性評価の結果

5. まとめ

本稿では波形接続型感情音声合成のための DNN を用いた感情制御法について述べた．本手法ではデータ拡張により，所望の感情制御則に従って学習データを生成し，これらを学習に用いることで DNN に制御則を埋め込む．評価において，データ拡張によって所望の制御則を持った DNN が学習できることを確認した．また主観評価において，従来法よりも適切な感情制御が行えることを確認した．

謝辞 本研究は公益財団法人東京都中小企業振興公社平成 29 年度中小企業経営・技術活用化助成事業の支援により実施した．

参考文献

[1] P. Taylor, “Text-to-speech synthesis,” *Cambridge University Press New York*, 2009.
[2] M. Schröder, “Emotional speech synthesis: A review,” *Proc. EUROSPEECH 2001*, vol.1, pp.561-564, Sept. 2001
[3] K. Tokuda, T. Kobayashi and S. Imai, “Speech parameter generation from HMM using dynamic features,” *Proc. ICASSP*, vol. 1, pp. 660–663, May 1995.
[4] H. Zen, K. Tokuda, T. Masuko, T. Kobayashi and T. Kitamura, “A hidden semi-Markov model-based speech synthesis system,“ *IEICE Trans. Inf. & Syst.*, vol. E90-D, no. 5, pp. 825-834, May 2007.
[5] H. Zen, A. Senior and Mike Schuster, “Statistical parametric speech synthesis using deep neural networks,” *Proc. ICASSP*, pp. 7962–7966, May 2013.
[6] J. Yamagishi, K. Onishi, K. Masuko and T. Kobayashi, “Acoustic modeling of speaking style and emotional expressions in HMM-based speech synthesis,” *IEICE Trans. Inf. & Syst.*, vol. E88-D, no. 3, pp. 502-509, Mar. 2005.
[7] M. Tachibana, J. Yamagishi, T. Masuko and T. Kobayashi, “Speech synthesis with various emotional expressions and speaking styles by style interporation and morphing,” *IEICE Trans. Inf. & Syst.*, vol. E88-D, no. 3, pp. 2484-2491, Nov. 2005.
[8] J. Parker, Y. Stylianou and R. Cipolla, “Adaptation of an Expressive Single Speaker Deep Neural Network Speech Synthesis System,” *Proc. ICASSP2018*, pp. 5309–5313, Apr. 2018.
[9] H. Yang, W. Zhang and P. Zhi, “A DNN-based emotional speech synthesis by speaker adaptation,” *Proc. APSIPA2018*, pp. 633–637, Nov. 2018.

ⓒ 2019 Information Processing Society of Japan

[10] Y. Ohtani, Y Nasu, M. Morita and M. Akamine, "Emotional transplant in statistical speech synthesis based on emotion additive model," *Proc. INTERSPEECH2015*, pp. 274-278, Sept. 2015.
[11] K. Inoue, S. Hara, M. Abe, N. Hojo and Yusuke Ijima, "An Investigation to Transplant Emotional Expressions in DNN-based TTS Synthesis," *Proc. APSIPA2017*, pp. 1253–1258, Dec. 2017.
[12] M. Abe, S. Nakamura, K. Shikano and H. Kuwabara, "Voice conversion through vector quantization," *Journal of the Acoustic Society of Japan (E)*, vol. 1, no. 2, pp. 71–76, 1990.
[13] Y. Stylianou, O. Cappé and E. Moulines, "Continuous probabilistic transform for voice conversion," *IEEE Trans. Speech and Audio Processing*, vol. 6, no. 2, pp. 131–142, 1998.
[14] T. Toda, A. W. Black and K. Tokuda, "Voice conversion based on maximum likelihood estimation of spectral parameter trajectory," *IEEE Trans. on ASLP*, vol. 15, no. 8, pp. 2222–2235, Nov. 2007.
[15] Y. Saito, S. Takamichi and H. Saruwatari, "Voice conversion using input-to-output highway networks," *IEICE Trans. Inf. & Syst.*, vol. E100D, no. 8, pp. 1925–1928, 2017.
[16] K. Kobayashi, T. Toda, G. Neubig, S. Sakti and S. Nakamura, "Statistical singing voice conversion with direct waveform modification based on the spectrum differential," *proc. INTERSPEECH*, pp. 2514–2518, Sept. 2014.
[17] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Proc. NIPS'12*, vol. 1, pp. 1097–1105, Dec. 2012.
[18] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Communication*, vol. 9, pp. 453–467, 1990.
[19] M. Morise, F. Yokonori and K. Ozawa, "WORLD: A vocoder-based high-quality speech synthesis system for real-time applications," *IEICE Trans. Inf. & syst.*, Vol. E99-D, no. 7, pp.1877–1884, 2016.
[20] SPTK Working Group, Speech processing toolkit, *http://sp-tk.sourceforge.net/*
[21] S. Imai, et al., "Mel Log Spectrum Approximation (MLSA) filter for speech synthesis," *Electronics and Communications in Japan (Part I Communications)*, vol. 66, issue 2 pp.10–18, 1983.
[22] F. Chollet, *et al.*, Keras, *https://keras.io*, 2015.
[23] M. Abadi, *et al.*, TensorFlow, *https://tensorflow.org*, 2015.
[24] D. P. Kingma, and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980*, 2014.
[25] I. J. Goodfellow, et al., "Generative Adversarial Networks,", *arXiv:1406.2661*, 2014.
[26] Y. Sato et al., "Statistical parametric speech synthesis incorporating generative adversarial networks," *IEEE/ACM Trans. ASLP*, vol. 26, issue 1, pp. 84–96, 2018.
[27] X. Mao, et al., "Least squares generative adversarial networks," arxiv:1611.04076, 2017

ⓒ 2019 Information Processing Society of Japan