# 音声からの感情・態度の理解

Recognition of Emotional and Attitudinal Information Conveyed by Speech

森　大毅

Abstract

音声は，言語情報だけではなく，話者の感情や態度などの情報を伝達する．このような情報を用いることで，人間と機械のコミュニケーションをもっと自然で豊かなものにすることが望まれる．本稿では，音声からの感情や態度などの認識技術の現状，及び音声インタフェースへの応用について解説するとともに，音声からの感情・態度の理解に向けた今後の研究を展望する．

キーワード：感情，パラ言語情報，音響的特徴，コーパス

## 1. はじめに

音声をテキストに書き起こすと，音声が本来持っていた情報は一部欠落するが，その情報の中には，音声コミュニケーションの本質と言ってもよい重要な要素が含まれている[(1)]．一般的な音声対話システムでは，ユーザ発話からの言語理解の前段に音声認識モジュールが置かれるが，音声認識の目的は音声のテキスト化であるから，テキストにできない情報，すなわちパラ言語情報はここで欠落する．このような情報を捨てることなくすくい上げ，人間と機械のコミュニケーションをより自然で豊かなものにすることは当然に望まれるが，それはいまだ十分に達成されていない．

本稿は，音声コミュニケーションの重要な要素である感情及びその周辺に焦点を当て，音声からの認識技術及びその応用の現状を解説する．

この種の研究には，(a)感情の認識そのものを目的とするものと，(b)感情などに関連するパラ言語的特徴を利用して音声インタフェースを改善・高度化することを目的とするものがある．(a)の研究には具体的な応用を想定していないものも多く，この点で(a)と(b)の間には感情などを認識することの意味や意義に対する考え方の隔たりがあるように思われる．(a)を「役に立つ」ものとするためには，このギャップを埋める何かを発見しなければならない．その一助となることを願って，本稿ではまず(a)の問題設定のために必要な概念及びその記述について述べた後に，5.で(a)の，次いで，6.で(b)の研究の具体例を取り上げる．音声からの感情・態度の理解に向けた今後の展望は7.で考察する．

## 2. 感情・態度の記述

### 2.1 基本感情

心理学における感情の理論には，主に二つの考え方がある．一つはカテゴリカルな感情を前提とする基本感情説，もう一つは感情カテゴリーを認めず連続的な感情を前提とする次元説である．

Paul Ekmanらダーウィンの流れをくむ心理学の学派は，普遍的で文化に依存しない感情（＝基本感情）の存在を主張し続けている．基本感情の中でも，「怒り」「喜び」「悲しみ」「驚き」「恐れ」「嫌悪」は，これまでの感情研究の中で繰り返し取り上げられ，しばしば6大感情と呼ばれる．「6大」というネーミングは，感情がそれらだけではないことを意味するのだが，音声からの感情認識の研究には，感情はこれらのうちのどれかであるという仮定に基づいたものも多い．

基本感情は感情の全てではないので，基本感情カテゴリーのみから成る感情語セットを用いて実際のコミュニケーション場面における感情を適切に記述するのは不可

森　大毅　正員　宇都宮大学学術院
Hiroki MORI, Member (Academic Association, Utsunomiya University, Utsunomiya-shi, 321-8585 Japan).
電子情報通信学会誌　Vol.101 No.9 pp.902-907 2018年9月

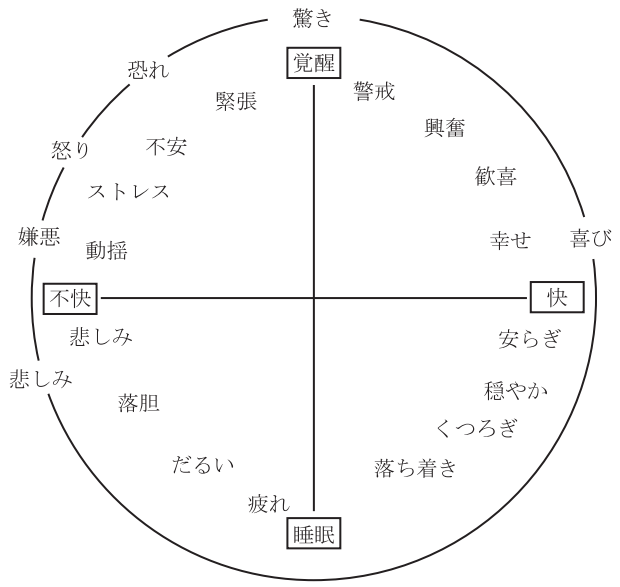


**図 1　感情の円環モデル**[3]　横軸が快―不快，縦軸が覚醒―睡眠の次元を表す．

能である．Cowie らは，よく生じる感情関連状態を含めたコンパクトな感情語セットを提案している[2]．

### 2.2　感情次元

次元説では，感情のカテゴリーは感情空間における布置を示すだけであり，重要なのはその空間を定義する次元であると考える．Russell は快―不快と覚醒―睡眠の二次元による円環モデル（図 1）を提唱した．快―不快は valence（感情価）と呼ばれることもあり，正の感情価は快に，負の感情価は不快に対応する．pleasantness（快度）または evaluation（評価）と呼ばれることもある．また，覚醒―睡眠は arousal（覚醒度）または activation（活性）とも呼ばれる．これらの二次元は，感情の基本的な次元として広く受け入れられている．また，第 3 の次元として dominance（支配性）を加えることも多い．

### 2.3　態度

心理学でいう態度とは，人の社会的行動に影響を及ぼす，ある特定の対象に対する評価や好悪のことである．

音声が伝達する態度を網羅的にカテゴリー化することは難しいと考えられ，相互行為上重要だと考えられる特定の態度に着目して記述することが一般的である．態度としてよく取り上げられる現象には，丁寧さ（politeness）や親しみ（friendliness）などがある．ほかに音声が伝達するパラ言語情報として「感心」「意外」「不満」などが取り上げられることがあるが，これらは感情と態度の双方に関連した概念であると考えられる．

感情次元と同様に，態度の側面を多元的に記述できる可能性もある．「肯定的態度」「否定的態度」の別は，その中でも最も基本的な対立軸となる[4]~[8]．

## 3. コ　ー　パ　ス

### 3.1　演技音声のコーパス

感情や態度を認識するためには，それらが音声の性質に与える影響をモデル化する必要がある．モデル構築のためには，感情や態度が表出された音声を何らかの方法で収録し，感情や態度のラベルを付与したコーパスが必要となる．このような音声の収録方法は，指示した感情を演技させて音声を収録する方法と，何も指示を与えず自然に生じたインタラクションにおける音声（自発音声）を収録する方法とに大別される．基本感情を研究の対象とする場合には，演技音声を収録することが多い．

感情と音声の研究史における中興の祖 Klaus Scherer とそのグループは，主に基本感情を研究の対象としていた．現在に至ってもなお，多くの感情認識研究で演技音声のコーパスが利用されている．Emo-DB[9]は，声優が指示された文を感情を込めて読み上げた，代表的な演技音声のデータセットである．SAVEE Database[10]も基本感情の演技音声を収録したコーパスである．IEMOCAP Database[11]も演技音声のコーパスであるが，より真正な感情表出に近いデータを得るため，二人の俳優に台本どおりに演じさせるセッションに加え，各感情に対して用意されたシナリオに従って即興で演技を行うセッションを実施している．

### 3.2　自発音声コーパス

音声対話システムとの対話，または人間同士の対話から感情を認識する場面では，自発音声からの感情認識が必要になる．一方，俳優や声優の演技は，自発的な感情表出とは異なる性質を持っており[12]~[14]，演技音声のコーパスから学習したモデルによる自発音声の感情認識は困難である[15]．このため，自然な感情表出を収録したコーパスの必要性が叫ばれてきた[16]．よく利用されるコーパスには，Vera am Mittag（VAM）[17]，RECOLA[18]がある．感情ラベルが付与された日本語の自発音声対話コーパスとして現在利用できるものは，宇都宮大学パラ言語情報研究向け音声対話データベース（UUDB）[19]，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）[20]の 2 種類である．

### 3.3　音声への感情／態度ラベリング

演技音声のコーパスの場合，各発話に感情／態度ラベルを付与する過程は自明となるが，収録後の聴取実験により，実験者が意図した感情が表出されていることを確認しておく必要がある．

自発音声コーパスの場合，音声サンプルと，それが発話されたときの話者の感情／態度との対応付けは自明で

はなく，これをどのように記述するかは，最も重要な課題の一つである．UUDB では感情次元が，OGVC では感情カテゴリーが，記述のために用いられている．IEMOCAP は，感情次元と感情カテゴリーの両方のラベルを持っている．

自発音声への感情／態度ラベリングでは，ラベルの信頼性が問題となる．信頼性を担保するために，ラベラ内・ラベラ間一致度を見積もっておくことは重要である[19]．

感情／態度のラベリング作業は極めて高コストであり，自発音声コーパスを使った研究のボトルネックとなっている．このため，感情／態度ラベルを有する自発音声コーパスは一般に規模が小さい．大量データの確保のため複数のコーパスを集約することは一つの解決策であるが[15]，異種コーパスでは感情／態度の記述法が異なるため，何らかの共通化が必要である[21),(22]．

大規模コーパスの構築では，質の高いラベラによる作業を代替する人海戦術が現実的方策となり得る．クラウドソーシングでの作業者数とラベルの安定性に関する調査[23]や，自動ラベリングの信頼度を利用した能動学習によりクラウドソーシングの作業量を最適化する試み[24]が報告されている．

## 4. 感情／態度認識に有効な特徴の抽出

感情及び態度によって影響を受ける音声の音響的特徴には，F0（基本周波数；声の高さに関連する），強度，テンポやリズムといった韻律的特徴，喉頭音源由来の声質パラメータ（平均的 F0，平均的強度，揺らぎ，喉頭雑音，声門開放率，スペクトル傾斜）[25]，声道由来の声質パラメータ（ホルマント；声道の共鳴特性）が含まれる．Scherer のメタ分析[26]によれば，覚醒寄りの感情である喜び・恐れでは F0・強度・話速が増大しスペクトル傾斜が減少する傾向が，睡眠寄りの感情である悲しみではその逆の傾向が一致して報告されているが，嫌悪その他の感情については余り一致していない．また，覚醒―睡眠に比べ，それ以外の次元の音響関連量は余り明確ではない．

上記の音響的特徴の大部分は超分節的（幾つかの音素にまたがって現れること）であり，個々の分析フレームよりも広い範囲にわたる変化が重要である．例えば F0 については，単に高いか低いかといった特徴に加え，ある範囲（例えば発話）での変動の大きさが感情の知覚に影響を与える．近年は，複数フレームにわたる音響的特徴の変化を要約した統計量を求める際，音声学的知識に基づいて厳選するのではなく，最大・最小・レンジ・平均・四分位数・標準偏差など考えられる統計量を全て利用することが一般的になっている．音声に関する重要な国際会議の一つである Interspeech では 2009 年から感情をはじめとしたパラ言語情報の認識精度を競うコンテストが開催されており，2013 年の Computational Paralinguistics Challenge（ComParE）[27]における標準特徴は，強度関連の 4 種類，スペクトル関連の 54 種類，有声音源特性に関する 6 種類から成る計 64 種類の低次特徴量（LLD：Low Level Descriptor）の変動を発話単位に要約した計 6,373 の高次特徴量（functional）から構成されている．openSMILE[28]はこのような大規模特徴抽出に向くフリーソフトウェアであり，フレーム単位の LLD，及び複数フレームにわたる統計量（functional）を得ることができる（図 2）．

このようにして得られる特徴ベクトルは非常に高次元になるため，過学習による性能低下の恐れがある．この問題に対し，マルチカーネル学習[29]や敵対的自己符号化器[30]による次元削減が提案されている．また，従来の総当り式の高次特徴量に代わり，DNN による表現学習が盛んに研究されている[31)~(34]．

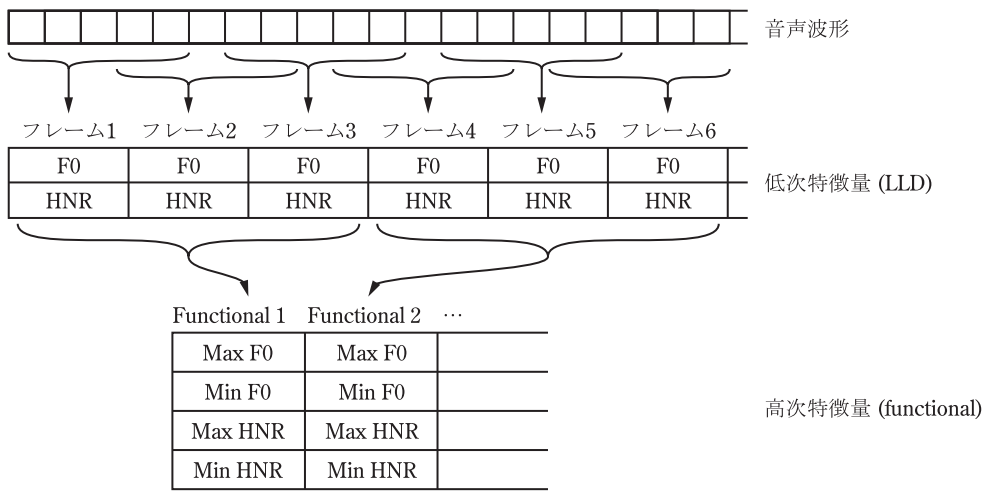


**図 2　LLD と高次特徴量の抽出**[28]　　LLD の変化を発話などの単位で要約し高次特徴量を得る．

発話全体の LLD から画一的に求められる高次特徴量は，感情や態度に関連する局在的な特徴（例えば句末音調など）を十分に反映することができない．これに対し，可変長の LLD を入力，長さ 1 の感情ラベルを出力とする注意機構（attention）付き符号器―複合器モデルにより，認識に有効な部分だけを利用する方法が提案されている[35),(36)]．

近年の DNN 音声認識では，従来の MFCC などの特徴抽出の代わりに，畳込みニューラルネットワーク（CNN）を利用して，特徴抽出も含めネットワーク全体を最適化することが行われている．この考えに基づき，Trigeorgis らは RECOLA データベースに対して CNN と LSTM によるエンドツーエンド感情認識ネットワークを学習し，前処理としての特徴抽出を一切行わず音声波形のみから arousal 及び valence を高精度で認識できることを示している[37]．

## 5. 感情・態度の識別

識別すべき感情・態度などが基本感情などのカテゴリーにより記述されている場合，その識別は分類問題または検出問題となる．感情分類問題は，$n$ 種類の中から一つの感情カテゴリーを「正解」として割り当てられた未知の音声サンプルに対し，その音響的特徴から「正解」の感情カテゴリーを推定する問題である．2013 年の ComParE ではプロ俳優が演技した 12 種類の感情分類が扱われた．ベースラインシステム[27]の精度は 40.9%UAR（データ数の偏りを平均化した再現率）と低く，2015 年の研究でも 42.4%UAR 程度にとどまっている[29]．OGVC のような自発音声の場合には精度は更に低くなる[21]．一方，コールセンターにおける顧客の怒り検出[38]，防犯を目的とした感情検出[39]のように，感情識別を検出問題として設定した場合には，感情カテゴリーを何種類用意すべきかといった問題は生じない．

次元により記述されている場合は，本来は快―不快，覚醒―睡眠などの程度を予測する回帰問題となるが，快 vs 不快（2 クラス）及び覚醒 vs 睡眠（2 クラス）またはそれらの組合せの分類問題として設定される[24),(27)]場合もある．肯定的―否定的態度の分類は，快―不快の分類に類似した問題となり，音響的特徴のみからでは容易ではない（文献(8)では精度 52.7%）．

近年は，RECOLA のように感情次元が時間的に連続に付与されているコーパスを対象に，発話単位での回帰ではなく，時間的に連続な感情追跡を問題として設定した研究が増えている．感情変化のダイナミクスをモデル化するために，GMM（混合ガウス分布）による特徴―次元マッピング[40]，双方向 LSTM[41),(42)]，カルマンフィルタ[43]が用いられている．

複数の感情次元[44]，複数レベルで離散化した感情次元[42]，演技 vs 自発[45]のように，相互に関連するタスクを同時に学習することでタスク間の関連性を活用して識別性能を向上させるマルチタスク学習も，近年よく使われる手法である．

## 6. 音声インタフェースにおけるパラ言語情報の利用

韻律や声質といった音声の特徴が伝達する感情や態度などのパラ言語情報を音声対話システムなどで利用しようとする研究は，認識そのものの研究に比べれば少ない．これは，3. で紹介したコーパスから学習可能なものと，現実のシステムが取り扱うべきパラ言語情報との間にギャップがあるため，既存コーパスの活用が難しいことが一因である．

対話システムへの応用の例としては，音声認識誤りにユーザが気付いたことを検知しシステムの修復行動に利用するもの[46]，ロボットからの提案に対するユーザの否定的態度を認識し提案を修正するもの[5]，ロボットの感情表出をユーザの音声から認識した感情に同調させることで共感を形成するもの[47]，教授システムからの質問に対する生徒の応答音声から自信のなさを検出し後続する教授戦略に反映させるもの[48]などが挙げられる．これらの研究には，2. で挙げた感情に比べより周辺的な現象を扱ったものが多く，また言語や表情など音声以外のモダリティを併用するものが多い．

## 7. お わ り に

音声からの感情・態度などを含むパラ言語情報の認識及び応用に関する現状を述べた．

今後の展望を論じる上で重要な論点は，コーパスについてである．深層学習技術の進展は，ImageNet（画像）や Audio Set（音）などの大規模データセットに支えられてきた．パラ言語情報についても大規模コーパスの開発が望まれるが，音声の収録法や感情／態度ラベルを付与する方法は今に至っても試行錯誤の段階である．世界的に見れば，研究の興味範囲は笑いやフィラーなどの社会的シグナル[27]，嘘[49]，ユーモア[50]など，より周辺的な現象に広がる傾向にあり，これら全てを単一の大規模コーパスでカバーするのは困難だと思われる．

感情認識のためのコーパス開発の問題を打破するための一つの可能性は，逆説的ではあるが，感情認識をやめてしまうことかもしれない．例えば，エージェントの発話の話速や強度をユーザ発話に同調させることでエージェントへの信頼感を向上させる試みがある[51]．この例のように，感情表出を含むユーザの振舞いに対して適応的に振舞いを変化させる機械の実現には，それらの対応関係のモデル化さえできればよく，感情やパラ言語情

報の認識は必ずしも要しないはずである．感情／態度ラベリングのボトルネックや，感情・態度記述の表現力不足の問題も自然に回避できる．

機械の感情認識精度を上げることが，果たして感情を理解する機械を実現することだろうか．

「ただいま」

今日，上司に怒られてしまったあなたの声は沈んでいる．そんなあなたに，何があったか話を聞いてくれ，そっと寄り添ってくれる．そのような振舞いに，あなたは自分の感情が理解されたと感じるのではないだろうか．

## 文　　献


（1） 森　大毅，前川喜久雄，粕谷英樹，音声は何を伝えているか―感情・パラ言語情報・個人性の音声科学―，コロナ社，東京，2014.

（2） R. Cowie, E. Douglas-Cowie, B. Apolloni, J. Taylor, A. Romano, and W. Fellenz, “What a neural net needs to know about emotion words,” in Computational Intelligence and Applications, N. Mastorakis, ed., pp. 109-114, World Scientific Engineering Society, 1999.

（3） J.A. Russell and G. Lemay, “Emotion concepts,” in Handbook of Emotions, 2nd edition, M. Lewis and J.M. Haviland-Jones, eds., pp. 491-503, Guilford Press, New York, 2000.

（4） D. Hillard, M. Ostendorf, and E. Shriberg, “Detection of agreement vs. disagreement in meetings : Training with unlabeled data,” Proc. HLT-NAACL 2003, pp. 34-36, 2003.

（5） 藤江真也，江尻　康，菊池英明，小林哲則，“肯定的／否定的発話態度の認識とその音声対話システムへの応用，”信学論(D-Ⅱ)，vol. J88-D-Ⅱ，no. 3, pp. 489-498, March 2005.

（6） G.A. Levow, V. Freeman, A. Hrynkevich, M. Ostendorf, R. Wright, J. Chan, Y. Luan, and T. Tran, “Recognition of stance strength and polarity in spontaneous speech,” Proc. 2014 IEEE Spoken Language Technology Workshop, pp. 236-241, 2014.

（7） L. Li, Z. Wu, M. Xu, H. Meng, and L. Cai, “Combining CNN and BLSTM to extract textual and acoustic features for recognizing stances in Mandarin ideological debate competition,” Proc. Interspeech 2016, pp. 1392-1396, 2016.

（8） G.A. Levow and R.A. Wright, “Exploring dynamic measures of stance in spoken interaction,” Proc. Interspeech 2017, pp. 1452-1456, 2017.

（9） F. Burkhardt, A. Paeschke, M. Rolfes, W.F. Sendlmeier, and B. Weiss, “A database of German emotional speech,” Proc. Interspeech 2005, pp. 3-6, 2005.

（10） S. Haq and P.J.B. Jackson, “Multimodal emotion recognition,” in Machine Audition : Principles, Algorithms and Systems, pp. 398-423, IGI Global, Hershey PA, 2010.

（11） C. Busso, M. Bulut, C.C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J.N. Chang, S. Lee, and S.S. Narayanan, “IEMOCAP : Interactive emotional dyadic motion capture database,” Lang. Resour. Eval., vol. 42, no. 4, pp. 335-359, 2008.

（12） K.R. Scherer, “Vocal communication of emotion : A review of research paradigms,” Speech Commun., vol. 40, no. 1-2, pp. 227-256, 2003.

（13） E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach, “Emotional speech : Towards a new generation of databases,” Speech Commun., vol. 40, no. 1-2, pp. 33-60, 2003.

（14） R. Cowie, “Perceiving emotion : Towards a realistic understanding of the task,” Philos. Trans. R. Soc. Lond. B, Biological Sciences, vol. 364, no. 1535, pp. 3515-3525, 2009.

（15） B. Schuller, B. Vlasenko, F. Eyben, M. Wöllmer, A. Stuhlsatz, A. Wendemuth, and G. Rigoll, “Cross-corpus acoustic emotion recognition : Variances and strategies,” IEEE Trans. Affective Computing, vol. 1, no. 2, pp. 119-131, 2010.

（16） B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, and S.S. Narayanan, “Paralinguistics in speech and language : State-of-the-art and the challenge,” Comput. Speech Lang., vol. 27, no. 1, pp. 4-39, 2013.

（17） M. Grimm, K. Kroschel, and S. Narayanan, “The Vera am Mittag German audio-visual emotional speech database,” Proc. ICME 2008, pp. 865-868, 2008.

（18） F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne, “Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions,” Proc. FG 2013, pp. 1-8, 2013.

（19） H. Mori, T. Satake, M. Nakamura, and H. Kasuya, “Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics,” Speech Commun., vol. 53, no. 1, pp. 36-50, 2011.

（20） Y. Arimoto, H. Kawatsu, S. Ohno, and H. Iida, “Naturalistic emotional speech collection paradigm with online game and its psychological and acoustical assessment,” Acoust. Sci. Technol., vol. 33, no. 6, pp. 359-369, 2012.

（21） H. Mori, A. Nagaoka, and Y. Arimoto, “Accuracy of automatic cross-corpus emotion labeling for conversational speech corpus commonization,” Proc. LREC 2016, pp. 4019-4023, 2016.

（22） 永岡　篤，森　大毅，有本泰子，“感情音声コーパス共通化のための新たな感情ラベル推定における既存感情ラベル併用の効果，”音響誌，vol. 73, no. 11, pp. 682-693, 2017.

（23） A. Burmania and C. Busso, “A stepwise analysis of aggregated crowdsourced labels describing multimodal emotional behaviors,” Proc. Interspeech 2017, pp. 152-156, 2017.

（24） S. Hantke, Z. Zhang, and B. Schuller, “Towards intelligent crowdsourcing for audio data annotation : Integrating active learning in the real world,” Proc. Interspeech 2017, pp. 3951-3955, 2017.

（25） R. Banse and K.R. Scherer, “Acoustic profiles in vocal emotion expression,” J. Pers. Soc. Psychol., vol. 70, no. 3, pp. 614-636, 1996.

（26） K.R. Scherer, “Vocal affect expression : a review and a model for future research,” Psychol. Bull., vol. 99, no. 2, pp. 143-165, 1986.

（27） B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, “The Interspeech 2013 computational paralinguistics challenge : Social signals, conflict, emotion, autism,” Proc. Interspeech 2013, pp. 148-152, 2013.

（28） F. Eyben, M. Wöllmer, and B. Schuller, “openSMILE : The Munich versatile and fast open-source audio feature extractor,” Proc. ACM Multimedia, pp. 1459-1462, 2010.

（29） X. Xu, J. Deng, W. Zheng, L. Zhao, and B. Schuller, “Dimensionality reduction for speech emotion features by multiscale kernels,” Proc. Interspeech 2015, pp. 1532-1536, 2015.

（30） S. Sahu, R. Gupta, G. Sivaraman, W. AbdAlmageed, and C. Espy-Wilson, “Adversarial auto-encoders for speech based emotion recognition,” Proc. Interspeech 2017, pp. 1243-1247, 2017.

（31） K. Han, D. Yu, and I. Tashev, “Speech emotion recognition using deep neural network and extreme learning machine,” Proc. Interspeech 2014, pp. 223-227, 2014.

（32） J. Lee and I. Tashev, “High-level feature representation using recurrent neural network for speech emotion recognition,” Proc. Interspeech 2015, pp. 1537-1540, 2015.

（33） M. Schmitt, F. Ringeval, and B. Schuller, “At the border of acoustics and linguistics : Bag-of-audio-words for the recognition of emotions in speech,” Proc. Interspeech 2016, pp. 495-499, 2016.

（34） S. Ghosh, E. Laksana, L.-P. Morency, and S. Scherer, “Representation learning for speech emotion recognition,” Proc. Interspeech 2016, pp. 3603-3607, 2016.

（35） C.W. Huang and S.S. Narayanan, “Attention assisted discovery of sub-utterance structure in speech emotion recognition,” Proc. Interspeech 2016, pp. 1387-1391, 2016.

（36） M. Neumann and N.T. Vu, “Attentive convolutional neural network based speech emotion recognition : A study on the impact of input features, signal length, and acted speech,” Proc. Interspeech 2017, pp. 1263-1267, 2017.

（37） G. Trigeorgis, F. Ringeval, R. Brueckner, E. Marchi, M.A. Nicolaou, B. Schuller, and S. Zafeiriou, “Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network,”

Proc. ICASSP 2016, pp. 5200-5204, 2016.
(38) D. Pappas, I. Androutsopoulos, and H. Papageorgiou, “Anger detection in call center dialogues,” Proc. CogInfoCom ’15, pp. 139-144, 2015.
(39) B. Schuller, M. Wimmer, D. Arsic, T. Moosmayr, and G. Rigoll, “Detection of security related affect and behaviour in passenger transport,” Proc. Interspeech 2008, pp. 265-268, 2008.
(40) H. Khaki and E. Erzin, “Continuous emotion tracking using total variability space,” Proc. Interspeech 2015, pp. 1299-1303, 2015.
(41) L. He, D. Jiang, L. Yang, E. Pei, P. Wu, and H. Sahli, “Multimodal affective dimension prediction using deep bidirectional long short-term memory recurrent neural networks,” Proc. AVEC 2015, pp. 73-80, 2015.
(42) D. Le, Z. Aldeneh, and E.M. Provost, “Discretized continuous speech emotion recognition with multi-task deep recurrent neural network,” Proc. Interspeech 2017, pp. 1108-1112, 2017.
(43) Z. Huang and J. Epps, “An investigation of emotion dynamics and Kalman filtering for speech-based emotion prediction,” Proc. Interspeech 2017, pp. 3301-3305, 2017.
(44) S. Parthasarathy and C. Busso, “Jointly predicting arousal, valence and dominance with multi-task learning,” Proc. Interspeech 2017, pp. 1103-1107, 2017.
(45) J. Kim, G. Englebienne, K.P. Truong, and V. Evers, “Towards speech emotion recognition “in the wild” using aggregated corpora and deep multi-task learning,” Proc. Interspeech 2017, pp. 1113-1117, 2017.
(46) D. Litman, J. Hirschberg, and M. Swerts, “Predicting user reactions to system error,” Proc. ACL ’01, pp. 370-377, 2001.
(47) F. Hegel, T. Spexard, B. Wrede, G. Horstmann, and T. Vogt, “Playing a different imitation game : Interaction with an empathic android robot,” Proc. Humanoids 2006, pp. 56-61, 2006.
(48) K. Forbes-Riley and D. Litman, “Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor,” Speech Commun., vol. 53, no. 9-10, pp. 1115-1136, 2011.
(49) B. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J.K. Burgoon, A. Baird, A. Elkins, Y. Zhang, E. Coutinho, and K. Evanini, “The Interspeech 2016 computational paralinguistics challenge : Deception, sincerity & native language,” Proc. Interspeech 2016, pp. 2001-2005, 2016.
(50) D. Bertero and P. Fung, “Deep learning of audio and language features for humor prediction,” Proc. LREC 2016, pp. 496-501, 2016.
(51) R. Levitan, Š. Beuš, R.H. Gálvez, A. Gravano, F. Savoretti, M. Trnka, A. Weise, and J. Hirschberg, “Implementing acoustic-prosodic entrainment in a conversational avatar,” Proc. Interspeech 2016, pp. 1166-1170, 2016.

（平成 30 年 4 月 23 日受付　平成 30 年 5 月 7 日最終受付）

**森　大毅**（正員）

平 5 東北大・工・通信卒．平 10 同大学院博士後期課程了．博士（工学）．現在，宇都宮大学術院（電気電子工学科）准教授．音声対話，音声合成，福祉情報工学の研究に従事．著書「音声は何を伝えているか―感情・パラ言語情報・個人性の音声科学」（共著，コロナ社）など．