M-31

球面 SOM を用いた音声に含まれる感情の可視化

A Study on Visualization of Emotion in Speech Using Spherical SOM

○三上和也[1]，関根好文[2]
*Kazuya Mikami[1], Yoshifumi Sekine[2]

Abstract: Recently, many researchers have studied about various robots. Especially the humanoid robot is expected as man's partner. Therefore, communication with people is important for a humanoid robot. For communication with people, it is important that a humanoid robot presumes people's emotions. In this paper, we discuss visualization with a smooth phase relation of the emotions included in the speech. As a result, we show that visualization of 7 emotions included in sound can be performed by using a sphere SOM.

１．まえがき

ロボットは家庭や仕事場など人間の生活環境で利用されつつあり，様々なロボットについて研究，開発が行なわれている．特にヒューマノイドロボットは家事や介護などの，人間のパートナーとしての役割が期待され，サービス分野を始めとする幅広い分野で活用していくことが望まれている．ロボットが人間の感情を理解できるならば，人間とのコミュニケーションがスムーズに行え，人間のパートナーになると考えられている[1].

我々は，人間のさまざまな生活環境下でロボットが人間との共存を可能とするためには，人間と協調することが必要である．ロボットが感情を推定することはコミュニケーションを行なうための重要な手段であると考え，自らコミュニケーションを行うロボットの実現を目的に研究を行っている.

人間が行うコミュニケーションの方法として，多くの場合用いられるものが音声によるコミュニケーションである．人間の音声には言葉の意味による情報伝達だけではなく，非言語的情報も伝達することが可能であり，感情を伝えることができる．また，人間の音声の性質は大きく分けて 3 つに分類できる．すなわち，音声の最小単位である音素や日本語の平仮名１つと同じ長さを持つモーラを表す音韻．音声の質を表わす音質，記述書記できない抑揚やリズム，音長など示し，感情を含む韻律である.

また，感情を含む韻律情報の特徴量として，ピッチ周波数，声の大きさ，母音の平均持続時間があげられる．母音の平均持続時間については，1 つの発話が感情表現の最小単位であるとして，その発話の時間を発話中のモーラ数で平均したものを用いる.

今回，音声に含まれる感情を，怒り，恐怖，驚き，嫌悪，幸福，悲しみの通常用いられる６感情と，これらを含まない平静の感情を加えた７感情とし，自律的に概念獲得を行うことができるニューラルネットワークである SOM（Self-Organizing Maps:自己組織化マップ）を用いて，音声に含まれる 7 つの感情の可視化について検討を行った.

２．本論

今回用いた SOM は 1981 年，T.Kohonen によって発表された大脳皮質の視覚野をモデル化したニューラルネットワークモデルの一種であり，教師なし学習ニューラルネットワークモデルである．SOM は多次元の入力データ群を類似度に応じて，2 次元平面等に配置されたニューロンのひとつが反応し，自立的に分類を獲得するニューラルネットワークモデルである．SOM は入力データ空間と競合層の 2 つをもち，入力データ空間と競合層のニューロン，すなわちノード i が，参照ベクトル m_i でそれぞれ結合している．入力データ空間から入力される入力ベクトル x は n 次元の要素をもち，競合層のあるノード i と結合している m_i も n 次元の要

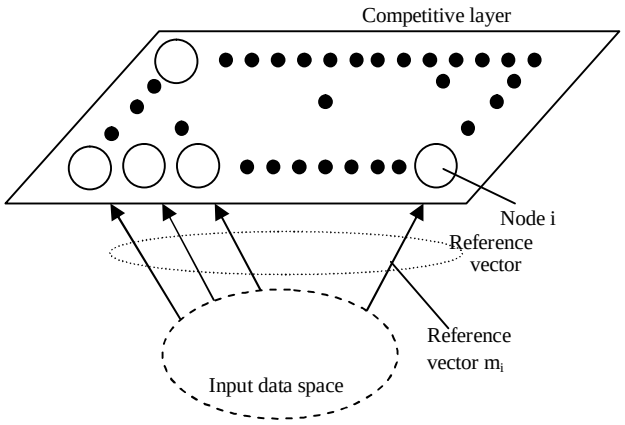


Figure 1. SOM.

１：日大理工・院・電子　２：日大理工・教員・子情

素を持つ．また m_i の初期値には無作為に選ばれた値が与えられる．自己組織化マップの学習は，x と m_i の差が一番小さくなるようなノードを選び，そのノードを BMU(Best matching unit)とする．BMU 近傍のノードは参照ベクトルを更新され、学習を行う．また，競合層を，多面体を用いて球面に近似することにより，2 次元平面では平面中央と端で起こる学習回数の偏りを解消できる[2]．球面に近似する多面体は，各頂点間の距離が等しい正多面体が望ましい．しかし，最も頂点数の多い正多面体の正二十面体は，頂点数が 12 個と少なく，SOM は競合層のノード数よりも少ない数のデータしか扱えないため，扱えるデータ数が限られる．そこで，正多面体をもとにし，球面上にほぼ均等に頂点を配置する格子構造を用いることとした．それぞれの頂点が入力データに対する特徴を示し，感情の位相関係がなめらかな球面 SOM の一例として，球面状に格子構造をもつ多面体を用い，各頂点をノードとし，各ノード間の類似度を辺で表現する構成とした．球面 SOM への入力する特徴量は，1 発話中の最大と最小ピッチ周波数，波形の最大振幅，発話時間の平均，最高ピッチ周波数と最低ピッチ周波数の差，ピッチ周波数の平均を用いた[3]．また，今回使用した音声サンプルは，

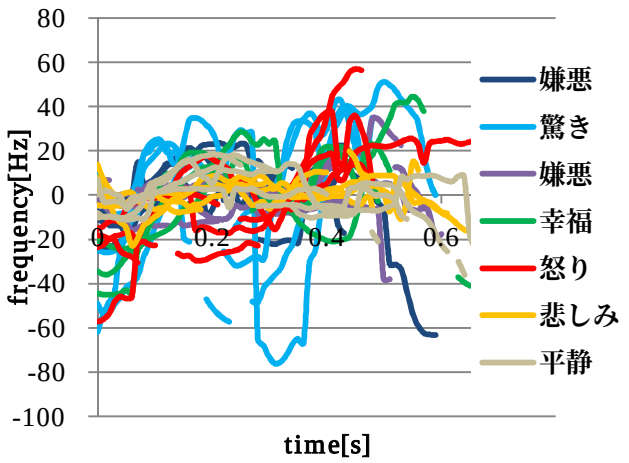
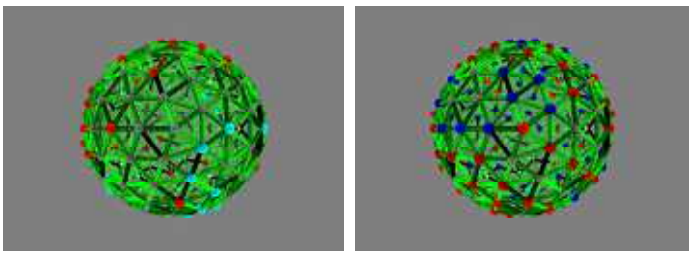


Figure 2.　Pitch frequency of 7 emotions.



(a)Viewpoint <0，0，3>　　(b) fuzzy c-means Emphasized cluster

Figure 3.　The simulation result of the sphere SOM.

2 句からなる文に対し，怒り，恐怖，驚き，嫌悪，幸福，悲しみ，平静をそれぞれ意識して発話したものを使用した．図 2 に 7 感情のピッチ周波数の平均値からの変化のグラフを示す．同図より，驚きや怒りはピッチ周波数の変化が大きく，また平静や悲しみは変化が少ない．図 3 に球面 SOM で可視化を行った結果の一例を示す．同図は，球面競合層を正二十面体を基とした三百二十面体とし，平静－怒りについて可視化を行った結果であり，三百二十面体の各頂点をノードとし，各辺の色の明度は各ノード間の類似度を表している．同図(a)は怒りの入力に反応するノードを赤色，平静の入力に反応するノードを青色で示す．この結果をファジィ c-means 法を用いてクラスタリングを行った．ファジィ c-means 法は複数のクラスタへ所属させる事がき，その所属の度合いをファジィ的に表現できるクラスタリング手法である．入力が怒りと平静の２つとしたので，所属されるクラスタを 2 つとする構成とした．クラスタリング結果より，片方に強く所属するノードをとりだしたものを同図(b)に示す．(a)と(b)の結果を比較すると一致する．このことは，可視化に対し球面 SOM が有効であることを示している．

３．まとめ

以上，自律的に概念獲得を行うことができるニューラルネットワークである SOM を用いて，音声に含まれる 7 つの感情の可視化について検討を行った．その結果，球面 SOM による可視化のシミュレーション結果とファジィ c-means 法による結果が一致したため，音声に含まれる７つ感情について可視化可能であり，可視化に対し球面 SOM が有効であることを明らかにした．

今後，複数の感情が同時に表現されることもあるため複合した感情について検討を行う予定である．

４．参考文献

[1] 原文雄:“人工感情-人とロボットの心の通うコミュニケーションの実現に向けて-”，日本機械学會誌 Vol.95, No.883, 508-512 , 1992.

[2] 高塚 正浩：球面 SOM のデータ構造と量子化誤差の考察およびインタラクティビティの向上，知能と情報 ： 日本知能情報ファジィ学会誌，Vol.19，No.6 pp.611-617 ,2007.

[3] 重永 實：「感情の判別からみた感情音声の特性」，電子情報通信学会論文誌，Vol.J83-A，No.6 pp.726-735 ,2000.