K-014

音声に含まれる感情情報の認識に関する研究

A study Recognition of the Feeling Information Included in Speech

横井 翔† 金子 正人‡ 武内 惇‡ 藤本 洋‡
Yokoi Shou Masato Kaneko Atsushi Takeuchi Hiroshi Fujimoto

1．はじめに

近年,コンピュータシステムの適用分野が拡大し，多くの人がコンピュータシステムを使用する機会が増加しており，様々な個性を持ったユーザに対して最適な環境を提供するHMI（Human Machine Interface）が必要であると考えられる．そこで我々は，ユーザの個性，感性を扱う手段として，音声に含まれる感情情報に着目し，ファジィ理論を利用した感情の抽出と認識を試みてきた[1]．ただ，そのファジィ理論を使った手法を用いると，１サンプルに対しては１感情のみの認識結果が得られるだけであった．実際の音声には，例えば悲しみと嫌々のように，複数の感情が含まれており，それらを含んだ認識結果を得られるようにしたい．

今回はニューラルネットワークの誤差逆伝播法[2]を用いた感情認識実験を行い，学習データについては複数の感情を認識することができた．

また，複数感情の認識結果を表す手段として認識距離という考え方を提案し適用したところ，結果を的確に表すことができたので，認識実験結果とあわせて報告する．

2．感情情報の抽出方法

2．1 発話実験

音声に含まれる感情情報の認識実験のためのサンプルを得るために発話実験を行う．発話実験ではあらかじめ定めた言葉（単文「あー，雪だ」）を成人男女に感情（「平静」「喜び」「悲しみ」「嫌々」「怒り」）を付加してそれぞれ発話してもらう．

2．2 聴取実験

音声サンプル「あー，雪だ」に対しその妥当性を確かめるため、またニューラルネットを用いた学習時の教師データとするために聴取実験を行った．聴取実験では最初に「平静」のサンプルを聞いてもらい，その後で各感情音声をランダムに聞いてもらう．各サンプルにつき３回の聴取によって被験者に感情を評価してもらう．これらの評価は感じられる感情の強さの印象を０（まったく感じない）から１（最も感じられる）の間のファジィ形式で複数回答可の条件で行う[1]．

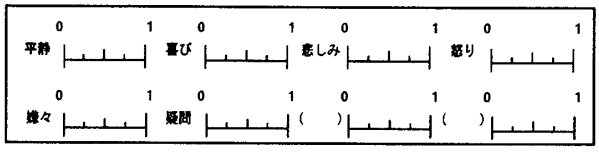

図１．聴取実験における感情評価記入欄

2．3 感情パラメータの解析

採取した音声データより，音響パラメータの解析を行う．単文（「あー，雪だ」：平静）の解析波形の例を図２に示す．感情パラメータとしてピッチ周波数の「最大ピッチ」，「最小ピッチ」，「平均ピッチ」，「始まりのピッチ」，「終わりのピッチ」，「声の大きさ」，「時間長」，「発声時間」の８個とした．

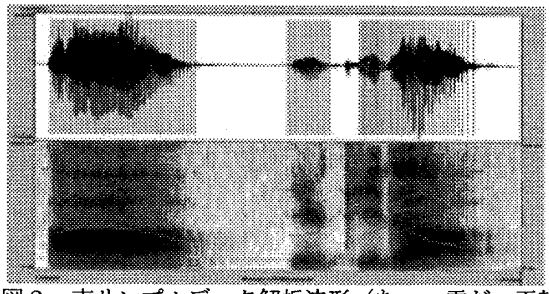

図２．声サンプルデータ解析波形（あー，雪だ：平静）

予備実験の結果から適切であると考えるニューラルネットワークの構成（入力層のユニット数８，第１中間層のユニット数を１６，第２中間層のユニット数を１４，出力層のユニット数６）でサンプルのパラメータ解析によって得た８次元のパラメータを入力層に設定し，聴取実験で得た聴取結果を教師データに設定し，学習を行う．

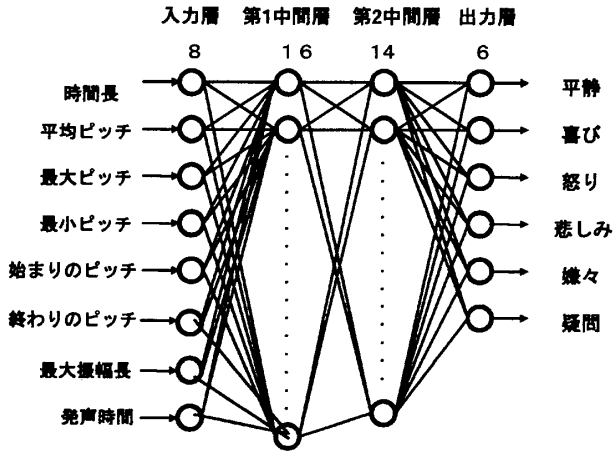

図３．学習と認識実験におけるニューラルネット構成図

2．4 認識実験

学習に用いるサンプル数は16人の５感情で90サンプルを学習させ，認識実験には学習データと同一者別データ及び未学習データを用いた．

学習データとは学習に用いた16人の音声データから無作為に選んだ３人分の音声データである．３人分の15サンプルの音声データを認識させその平均を結果として用いた．

† 日本大学大学院工学研究科情報工学専攻
‡ 日本大学工学部

同一者別データとは学習に用いた 16 人の音声データに含まれる一人から別の日に採取した音声データである．これを 3 日分採取し，15 サンプルの音声データを認識させその平均を結果として用いた．

未学習データとは学習に用いた 16 人の音声データ以外の 1 度も学習させたことのない音声データから 3 人分の 15 サンプルの音声データを認識させその平均を結果として用いた．

3．結果と考察

3．1　認識結果

学習データでは 9 割以上の認識結果を得ることができた．

ニューラルネットによる感情認識の特徴の一つに複数感情の認識可能という点が挙げられる．よってその中から有効な値を求めることが必要になってくる．しかし，表 1 のようにそれらの平均をとると，該当の感情の結果が思わしくなかった場合に違和感のある結果となってしまう場合がある．該当の感情だけを有効な値として用いると複数感情を認識可能というニューラルネットの特徴を生かしきれないという問題がある．

表 1．未学習データの認識結果

		平静	喜び	悲しみ	怒り	嫌々	疑問	平均認識率
平静	聴取結果	0.7222	0.0093	0.0000	0.0185	0.0000	0.0296	
	認識結果	0.6917	0.6509	0.0000	0.0361	0.0000	0.0000	
	認識率	97%	36%	100%	98%	100%	97%	88%
喜び	聴取結果	0.0093	0.7833	0.0000	0.0000	0.0000	0.0000	
	認識結果	0.0009	0.1102	0.0392	0.7770	0.2582	0.0000	
	認識率	90%	33%	96%	77%	74%	100%	71%
悲しみ	聴取結果	0.0000	0.0000	0.6278	0.0000	0.0389	0.2037	
	認識結果	0.0015	0.1258	0.0480	0.5510	0.2350	0.0000	
	認識率	100%	91%	47%	45%	86%	90%	72%
怒り	聴取結果	0.0000	0.0000	0.0000	0.5833	0.1574	0.0000	
	認識結果	0.4962	0.0052	0.0000	0.0548	0.0011	0.0000	
	認識率	50%	99%	100%	47%	84%	100%	80%
嫌々	聴取結果	0.0000	0.0000	0.0778	0.3204	0.4167	0.0000	
	認識結果	0.0074	0.0259	0.0207	0.5100	0.1775	0.0000	
	認識率	99%	97%	89%	41%	76%	100%	91%

3．2　認識距離の導入

そこでこれらの問題を解決し，有効な結果を表すために認識距離というものを提案する．この認識距離とは，表 2 に示すように聴取結果と認識結果の差をそれぞれ求めその差をすべて合計したものである．また，表 3 はそれらの認識距離を学習データ，同一者別データ，未学習データごとに平均をとったものである．

表 2．認識距離の例

		平静	喜び	悲しみ	怒り	嫌々	疑問	認識距離
平静	聴取結果	0.7222	0.0093	0.0000	0.0185	0.0000	0.0296	
	認識結果	0.6917	0.6509	0.0000	0.0361	0.0000	0.0000	
	差	0.0305	0.6416	0.0000	0.0176	0.0000	0.0296	0.7193
喜び	聴取結果	0.0093	0.7833	0.0000	0.0000	0.0000	0.0000	
	認識結果	0.0009	0.1102	0.0392	0.7770	0.2582	0.0000	
	差	0.0084	0.6731	0.0392	0.7770	0.2582	0.0000	1.7558
悲しみ	聴取結果	0.0000	0.0000	0.6278	0.0000	0.0389	0.2037	
	認識結果	0.0015	0.1258	0.0480	0.5510	0.2350	0.0000	
	差	0.0015	0.1258	0.5798	0.5510	0.1961	0.2037	1.6581
怒り	聴取結果	0.0000	0.0000	0.0000	0.5833	0.1574	0.0000	
	認識結果	0.4962	0.0052	0.0000	0.0548	0.0011	0.0000	
	差	0.4962	0.0052	0.0000	0.5285	0.1563	0.0000	1.1861
嫌々	聴取結果	0.0000	0.0000	0.0778	0.3204	0.4167	0.0000	
	認識結果	0.0074	0.0259	0.0207	0.5100	0.1775	0.0000	
	差	0.0074	0.0259	0.0571	0.1896	0.2392	0.0000	0.5192

表 3．5 感情の認識距離（平均）

	学習データ	同一者別データ	未学習データ
平静	0.1300	0.3367	0.7902
喜び	0.0251	0.3800	1.1093
悲しみ	0.0611	1.2270	1.4218
怒り	0.0614	1.0959	1.1985
嫌々	0.0370	0.9560	0.6934

図 4 は各認識距離の比較である．全体的に学習データ，同一者別データ，未学習データの順で認識距離が短い結果を得られた．平静と喜びの同一者別データは未学習に比べ認識距離が半減しているが、その他の感情についてはあまり良好な結果が得られなかった．しかし認識距離という考え方を提案し適用することで結果を的確に表すことができた．

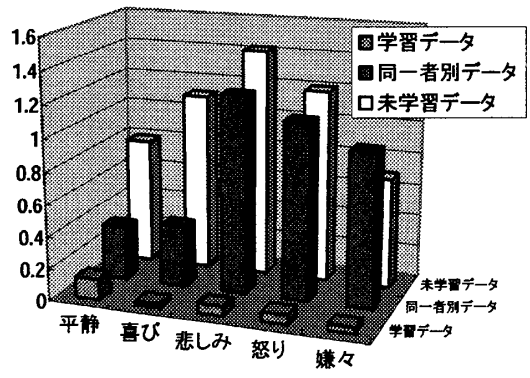

図 4．認識距離の比較

4.　終わりに

今回はニューラルネットワークの誤差逆伝播法を用いた感情認識実験を行い，学習データについては複数の感情を認識することができた．しかし未学習データについては良い認識結果を得ることはできなかった．

また，複数感情の認識結果を表す手段として，認識距離という考え方を提案し適用したところ，結果を的確に表すことができた．

今後の課題として，未学習データの認識も可能となるよう現在まだ少量であるサンプルデータを増加させる必要がある．また，サンプル数が少ない場合にも対応できるような新しい認識方法も検討していきたい．

参考文献

[1] 佐藤，遠藤，金子，武内，藤本：“ 音声に含まれる感情情報の認識に関する研究”，
信学技報 HIP 2001-60，pp.25-29，2001 年

[2] 甘利俊一，向殿政男：“ ニューロとファジィ” アドバンスドエレクトロニクスシリーズⅡ-1 P245