人工知能学会研究会資料
SIG-SLUD-A101-7(6/15)

話し言葉コーパスにおける感情タグの認識要因について

Identifying emotional elements in a speech corpus

坊農真弓[1,6] 影山良之[2,6] 谷口未希[3,6] 飯田朱美[4,6] ニック・キャンベル[5,6]
Mayumi Bono, Yoshiyuki Kageyama, Miki Taniguchi, Akemi Iida, Nick Campbell
mayu-b@kcn.ne.jp, yoshi-ka@is.aist-nara.ac.jp, t-miki@dab.hi-ho.ne.jp, akeiida@sfc.keio.ac.jp, nick@isd.atr.co.jp
[1] 神戸大学大学院総合人間科学研究科
Graduate School of Cultural Studies and Human Science, Kobe University
[2] 奈良先端科学技術大学院大学情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology
[3] 大阪大学大学院言語文化研究科
Graduate School of Language and Culture, Osaka University
[4] 慶応義塾大学SFC研究所
Keio Research Institute at SFC, Keio University
[5]ATR 先端情報科学研究部
ATR Information Sciences Division
[6]CREST, JST (Japan Science and Technology)*

Abstract: This paper describes a study of methods for identifying speaker's emotional state in a speech corpus. For some time now, psychologists have been classifying emotions through the use of spatial dimensions, using a circumplex model. The European Community's PHYSTA project recently merged psychological and linguistic analyses to develop a hybrid system capable of using information from facial expression and voice to recognize people's emotions. In spoken dialog, people can usually recognize the speaker's emotion, attitude, and intention through implicit paralinguistic cues, as well as from gesture and facial expresssion. This paper shows that people discriminate such emotional state information from three types of information in the spontaneous speech, namely, context, word meanings, and prosody. In this paper, we examine the role of these factors in the recognition of emotions from speech and report results from experiments using the FEELTRACE tracking system proposed by the PHYSTA project.

1 はじめに

我々は，対話を行なう際，通常，音声を媒介とする．韻律情報に何らかの話者の意図や感情，態度が表現されるのではないかという考察はこれまで心理学や言語学，音響音声学などの分野で多岐にわたってなされてきた．心理学において感情を連続的であるとする次元説を唱える Russell[1][2] は，感情を2次元で表そうとした．Russell は，すべての感情は，本人の提唱する「円環モデル」でまかなえるとした．それまでは，感情の因子は離散的に存在するという基本感情説を唱える Plutchik[3] などにより，感情を8種類の独立した因子として，空間布置が行なわれてきた．Russell の独自性は，感情は「快—不快」「覚醒—眠気」の2次元上で連続的に存在するものであるとしたところにある (鈴木 [4])．感情表出の音響的特徴を観察することで，今後の音声合成などの分野に利用しようという試みはこれまでにも Iida 他 [5] などでなされてきた．また，徳久他 [6] では，言語コーパスに感情要素を付与する研究もなされている．実際に感情の要素をこれらの他分野に用いる際，人間はどのような要因を参照し，感情認識を行なうのかといった問題点について触れる必要があるのではないだろうか．音声言語の特徴の一つは，線条性を持っているということである．時間と共に成立して

*科学技術振興事業団 CREST「高度メディア」研究領域「発話様式プロジェクト」，〒619-0288 京都府相楽郡精華町光台2丁目2番地 (株) 国際電気通信基礎技術研究所,http://www.isd.atr.co.jp/esp/

いく音声言語において，我々は，何から話者の感情や話者意図を認識するのであろうか．3つの要素が予想できる．それらは，文脈情報，語意情報，音声情報である．文脈情報とは，それまで話されていた事柄であり，知識として長期にわたって持ちえていると予想できる情報である．語意情報とは，文脈情報と相反する関係にあり，瞬間的に発話される発話内の言語情報を指す．音声情報とは，速度や韻律，インテンシティー，声質などの情報であり，語意情報と同じく，瞬間的な情報である．文脈情報を通時的情報とし，語意情報，音声情報を共時的情報として分類が可能である．感情の認識に影響を及ぼすであろう，文脈情報に関しては，Russell[7]において表情の研究を通し，順応水準理論を挙げ指摘している．ここで，Russellは文脈を，「表出者と表出者を取り囲む環境，人物像，状況，直前に何を言ったのか」などの要素を含めて取り扱っている．

今回は，PHYSTAのFEELTRACEを利用し，上記した事柄に関する実験を行い，考察を進める．

2 FEELTRACE

Cowie他[8]では，これまで，言語学や工学で，人間のコミュニケーションにおける様々な情報伝達の仕組みを明らかにするために多くの努力がなされてきたが，更にこれからは，心理学や言語学の理論を取り入れるべきであると述べる．このヨーロッパを中心として展開する感情を扱うPHYSTA[1]というプロジェクトでは，顔や声から人間の感情を認識するための実験ツールも提案されている．中でもFEELTRACEという実験ツールは，音声言語における感情認識のために開発されたツールであり，心理学の理論を多く取り入れ，Cowie他[9]では，そのツールの適応性に関しても様々な実験が行なわれている．このFEELTRACEは，今後音響分析をする際に，感情的と認識される部分を抽出する手がかりとすることを目的として考案されている．

2.1 FEELTRACEの円環構造

このツールは，上記したようなRussellなどの「円環モデル」を利用して作られている．しかし，機能に関しては，異なる部分も多い．FEELTRACEは，2本の座標軸からなり，2次元モデルと言えそうであるが，実際は，2本の軸の交点から，円の縁までに強さの調節が入り，ベクトルの向きだけを重要視する「円環モデル」とは明らかに異なる．また，X軸を「快—不快」から，「positive－negative」とし，「覚醒—睡眠」を「active－passive」と捉えなおしている．

[1] http://www.image.ece.ntua.gr/physta/

2.2 感情用語の布置

このFEELTRACEの内側と外側に布置された感情用語に関してもPHYSTA独自に考案されたものである．日本語においても感情用語をこの2次元のどの位置に置くべきかという議論は，菊谷他[10]などで多くなされてきた．今回の観察では，PHYSTAの提案している感情用語の配置を変えず，日本語訳をし，用いる．

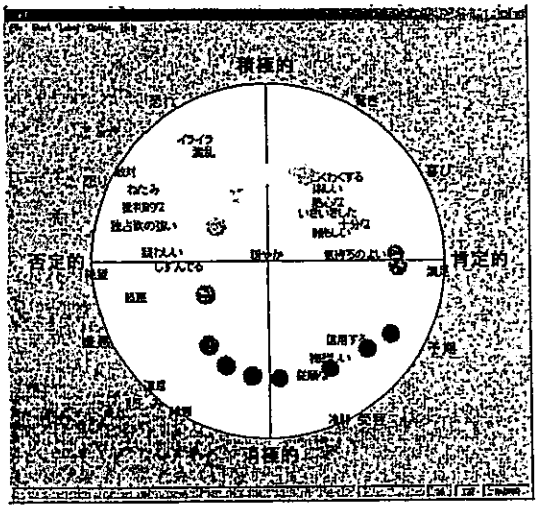

図1: FEELTRACE

3 観察方法

3.1 被験者

被験者は大学生，大学院生男性3名，女性14名とした．

3.2 刺激と装置

刺激は2種ある．一つは，発話者が意図的に「悲しみ」「喜び」「怒り」の3種類の感情を読み分け，防音室において，コンデンサマイク(SONY C-355)で収録したものである．今回は，感情用語を含まない，「オランダに生まれたゴッホは，伝道師，画商などの職を経て画家を志した」という文章を，刺激として採用した．この刺激は，個人レベルでこの発話の音声的特徴から3種の感情を聞き分け，ツール構造を理解しているか知るために使用した．この音声を本稿では，メジャー(測定)刺激と呼ぶ．もう一つの刺激が本稿の本刺激である．男性と女性の自然対話である．男性側にピンマイク(SONY-ECM77B)を口元から10cmの位置に装着させ収録した．女性側からのインタビュー形式の対話であり，女性のあいづち的発話も聞き取れる程度に収録してある．ここでは，30分56秒収録し，この中から，男性が発話権を取り，様々な話題に関して立て続

けに話している箇所を4分間抜粋した．この音声をここでは，コンテクスト刺激と呼ぶ．この刺激内の発話を区切り，ランダムに並べ替えた刺激も採用した．このとき，発話間に3.3秒ずつポーズを置いた．これをランダム刺激と呼ぶ．ランダム刺激の総時間は4分50秒である．

刺激の提示装置はパソコン (SONY VAIO PCG-XR1F/BP) を用いた．画面上には図1のようなFEELTRACEと同じ機能を持つ日本語で感情用語が書き込まれたものを使用した．

3.3 手続き

まず，ツールの説明として，画面上に映し出されている2次元のFEELLTRACEに関して説明を行なった．その際，マウスを実際に動かすように指示し，「積極的—消極的」「肯定的—否定的」の軸を説明し，円の中心から縁までの間で度合いを示す必要があることを教示した．また，今回は，ツール内に用いられている色に関しては，ツール使用上の援助的役割であるとし，感情用語と色彩に関連があるとは指示しなかった．まず始めに，トレーニングとして男女の対話を60秒聴かせ，画面上のツールを用い，男性の感情をトレースさせた (今回トレースさせる音声はすべて2回流し，1度目は聴くだけで，2度目にツール上でトレースをするよう，指示した)．次に，3種の感情を「喜び」「悲しみ」「怒り」の順序で並べた音声を聴かせ，その順序を説明し被験者に意識させ再度聴かせた．そして，音声を「悲しみ」「怒り」「喜び」の順序に変えたものを聴かせ，実際にトレースをさせ，ツールの使用や感情認識への判断の測定を行なった．本刺激である男女対話をランダム刺激，コンテクスト刺激という順序で聴かせ，男性の気持ちを画面上でトレースさせた．

4 結果と考察

4.1 個人差

図2は，メジャー刺激のトレース結果である．図2のように，(左下), (左上), (右上) の順序でトレース可能であった被験者は，被験者AからLまでの12名であった．この12名を実験前のトレーニングから，ツールの使用法を理解し，X軸（肯定的—否定的），Y軸（積極的—消極的）を理解して実験に望んだものとして，今回の分析対象とする．残る5名は図2のようなトレースを行わなかった．よって，今回は，この5名のトレース結果を分析の対象外とする．

図3,4,5は，「悲しみ」(左下)，「怒り」(左上)，「喜び」(右上) の各方向の最大値における個人間分布を示

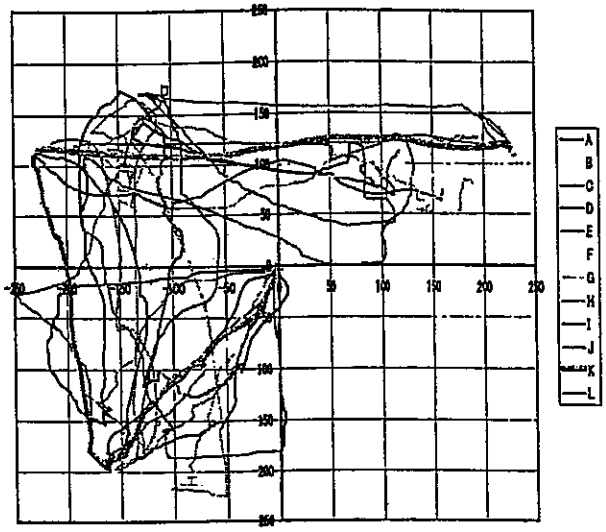

図2: メジャー刺激トレース結果（採用例）

している．ここでは，X軸とY軸の交点から，円の縁である，最大値までの感情の度合いの調節に，どこまで，個人間でばらつきが見られるかを観察した．この観察から，トレースする際，円内での度合い調節は，個人間でばらつくことが明確である．ツールに対する認識，感情認識などの個人間の程度差が原因であると考えられる．

各感情のトレースの最大値分布の揺れにおける程度は異なる．これらのトレースの最大値分布は，感情用語の布置が大きく関係してくるであろう．実験後の被験者に対する自由記述アンケートやインタビューにおいて，X軸とY軸よりもツール内にかかれている感情用語を意識したという意見もあった．今回はPHYSTAの挙げた感情用語を日本語訳して用いたが，被験者が日本人であることからも，今後は文化差なども考慮し日本語における感情用語の布置を考える必要がある．

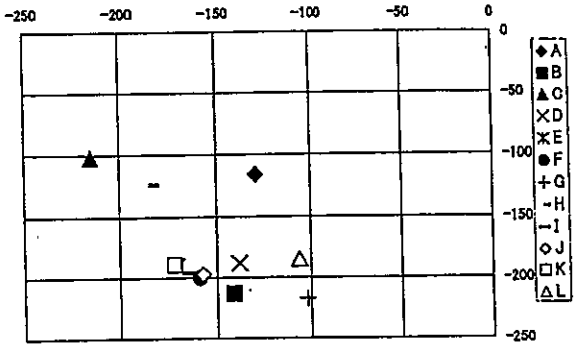

図3: メジャー刺激（悲しみ）最大値分布

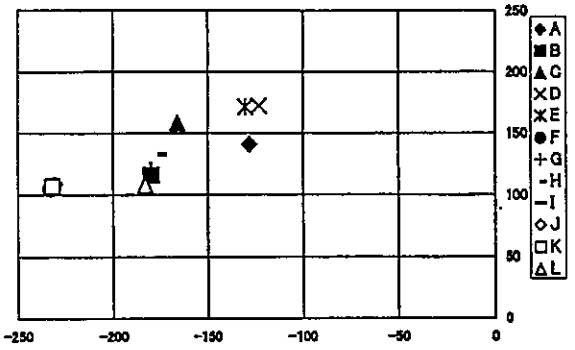

図 4: メジャー刺激（怒り）最大値分布

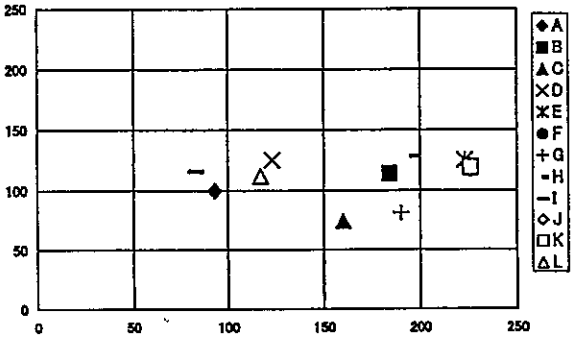

図 5: メジャー刺激（喜び）最大値分布

4.2 感情表現の音声特徴

今回メジャー刺激に用いた 3 種の感情音声は，上述したように，テキストは同じで，3 種の感情に読み分けられたものである．よって，被験者は，音声特徴から感情を認識し分けていると予測可能である．発話速度，インテンシティー，ピッチなどの様々な要因が考えられるが，宇津木 [11] の研究を踏まえ，基本周波数を以下の表 1 において比較する．

宇津木 [11] では，「怒り」は基本周波数での表出が抑制されるものとされないものがあると指摘し，基本周波数からのみの観察では不十分であると述べる．今回の「怒り」の平均周波数は宇津木 [11] の抑制されない「怒り」の平均周波数と一致する．「悲しみ」においてもこれまでの研究に比べ，周波数は高い．「喜び」などの快感情は不快感情に比べ，研究が少ない傾向にある．今後は，基本周波数以外の声質や速度などの観察を行い，聞き手の認識の状況を分析する必要がある．

4.3 文脈情報と語意情報

本実験では，コンテクスト刺激とランダム刺激を用いた．この目的は，発話をランダムに並べ替えることにより，文脈要素を減少させ，感情認識には文脈情報が関係するのか，それとも語意情報，音声情報が関

表 1: 感情 3 種の基本周波数

	悲しみ	怒り	喜び
最大値 [Hz]	285.0	307.0	313.0
最小値 [Hz]	152.0	148.0	160.0
平均値 [Hz]	206.6	210.3	240.6
標準偏差	33.1	34.1	44.2

係するのかをみることであった．しかし，今回は，個人間での感情認識の程度差やツールの使用における個人差が大きいため，発話間という小さい区切りの中には，感情認識の傾向を観察するのは不可能であった．以下の図 6 は，コンテクスト刺激における X 軸のトレース結果であり，図 7 は，ランダム刺激での X 軸のトレース結果である．ランダム刺激では，発話の順序を変えることで，話題転換構造を壊してしまい，話題転換から認識する話者意図のトレース一致がランダム刺激の結果から観察されなかったと考えられる．今回は，話題転換構造からの認識を観察するために，コンテクスト刺激の話題転換に関して分析を行なう．

図 6，図 7 を視覚的に見て観察できることは，コンテクスト刺激のほうが，感情認識の一致が比較的よく観察できるということである．図 6 のコンテクスト刺激の 58 秒から 78 秒までの付近で上昇傾向がある．また，79 秒から 143 秒付近では，X 軸の値が 0 から 100 の辺りに被験者が停滞する．また，144 秒から 188 秒辺りに下降傾向にある．このことに関して，話題転換の概念を利用し，文脈情報からの認識影響を考察する．

表 2: コンテクスト刺激の話題転換構造

時間 [sec]	話題		認識
0.0-54.9	自然言語収集		
54.9-77.9	サークル紹介		上昇
77.9-139.2	サークル主催者と 大学ネットワーク		停滞
77.9-205.0	サークル活動		下降
205.0-		ホームページ サーバ	
	(サークル)	(コンピュータ)	

表 2 のように今回のコンテクスト刺激の話題は，大きく分けると 2 つである．始めから，54.9 秒まで続く「自然言語収集」がまず一つである．次は，54.9 秒から最後まで続く「サークルとコンピュータ」に関する話題である．後者に関しては，大学時代のサークルと現

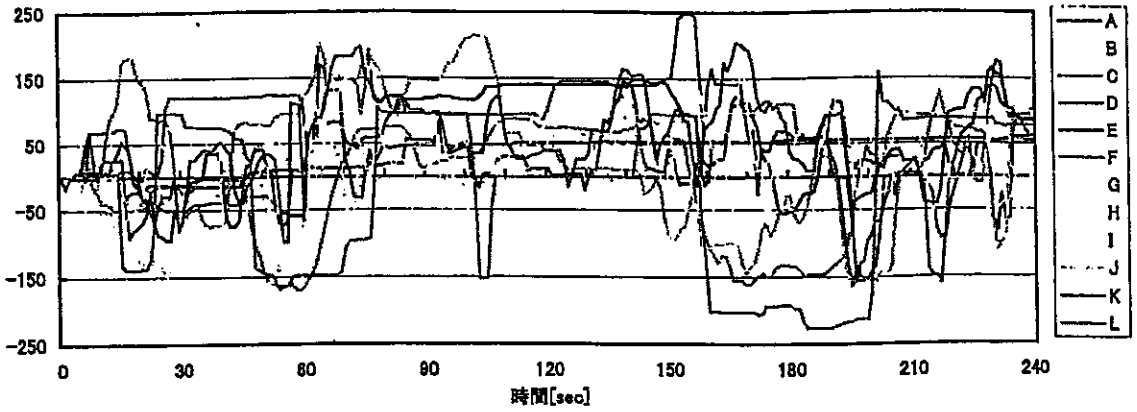

図 6: コンテクスト刺激のX軸のトレース

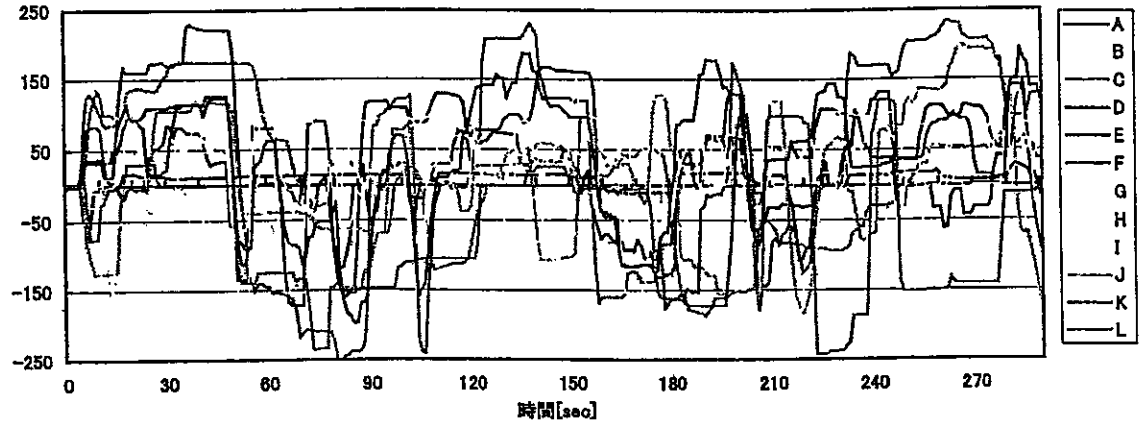

図 7: ランダム刺激のX軸のトレース

在使用しているサーバーに関して関わりを説明している．その話題の中で，更に細かく，「サークル」と「コンピュータ」の話題が交錯しながら会話が進められる．ここで，上述したX軸における被験者の認識が上昇傾向にある箇所の話題転換構造に注目する．この上昇部分では，男性話者は，女性話者の誘導により，話題を「自然言語収集」から転換させ，大学時代のサークルの話をし始める．これにより，12名の被験者は，男性の感情を上昇傾向にトレースしたと考えられる．次に認識が停滞した79秒から143秒付近では，男性の発話は，コンピュータと大学時代のサークルを結びつけるために説明的な発話を行なう．よって，ここは，X軸の値が0から100の間で停滞すると考えられる．次に，認識は下降傾向に移るのであるが，ここでは，男性話者は，再びサークル自体に関して触れ，冗談でサークルを批判する．このような話題転換が起きたために，被験者はトレースをマイナス方向へ移行させたと考えられる．今回のコンテクスト刺激の音声対話は，比較的安定した対話で，音声的に際立って感情表出する箇所は少なかった．よって，被験者は，音声から見た男性話者の感情よりもむしろ文脈の中での話題転換から話者意図をトレースした可能性も高い．今後検討を要する．

4.4 音声情報

コンテクスト刺激でのX軸の値において，8人以上のトレースの一致が見られた箇所は，6箇所あった．それらの箇所のうち4箇所は，トレースの寸前（刺激から反応までの遅れを考慮）に聴覚印象的に特徴のある音声があった．表3に挙げる．

音声Aの「あ」は明らかに感動詞で「驚き」の機能を持つ（坊農[12]）．4.3で触れたように，ここでまず，話題転換が起こり，感情認識も上昇傾向で被験者間に一致が見られる．この音声情報を合図に被験者が話題転換を意識し，男性話者の感情が変化したと認識した可能性がある．音声Bはつぶやくような音声で，これに関してはピッチ抽出が不可能であった．しかし，パワーの弱い音声に対して特定の感情を当てはめる傾向にあるという坊農[12]の指摘に一致する．音声Cは，笑い

表 3: X 座標値が 8 名以上一致した箇所

音声	時間	発話内容	x 座標
A	59.0-59.6	あっ，そう	9-13
B	172.6-173.2	そう，あほやねん	53
C	197.3-199.0	年取ったしね (笑)	56
D	210.0-212.2	うーん， 昔は楽しかった	56

声を含んだ発話であり，パラ言語的な要素から感情を認識したと考えることが可能である．音声 D は，F0 の値も低く，特定の音声表現を用いた発話である．これらの音声的特徴を持つ発話に対して，X 軸のトレースに一致が見られたことは，感情認識において，音声情報の援助が大きく影響することを意味する．

5 おわりに

これまで，文脈情報，語意情報，音声情報のどの要因が感情認識に最も影響を与えるのかを考察してきたが，どの要素が最も影響を与えるのか特定すること以上にそれらの 3 要因がどのように影響しあっているのかを観察することが今後の課題であろう．今回の実験では実際の自然対話を刺激として用いることにより，結果として被験者に感情認識の一致箇所があり，話題構造や音声特徴などと確実に関わることが観察可能であった．話し言葉のコーパスに感情要素をタグ付けする際に，これらの 3 要素のうち何に着目するかで，感情の評価は異なるのである．本研究は感情評価の要因の複雑さを更に明確に示す結果となった．今後は，話者の心理状況などを他者が認識する場合の要素を音響分析的な手法を用い，特定していく必要があるであろう．

謝辞

本稿執筆にあたり，心理学に関して同志社大学の鈴木直人教授，ならびに同志社大学心理学研究室のみなさまに有益なコメントをいただきました．また，千葉大学の伝康晴助教授には本研究に関して様々なコメントをいただきました．ここで，深く感謝いたします．実験において，協力していただいた方々にもこの場をお借りしてお礼申し上げます．

参考文献

[1] Russell, J.A. 1980 A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161-1178

[2] Russell, J.A. 1997 How shall an emotion be called? In Plutchik, R. & Contc, H. (Eds.) *Circumplex Model of Personality and Emotions.* Washington; APA

[3] Plutchik, R. 1960 The Multifactor-Analytic Theory of Emotion. *The Journal of Psychology*, 50, 153-171

[4] 鈴木直人 (印刷中)『感情心理学への招待』サイエンス社

[5] Iida, A., Campbell, N., Iga, S., Higuchi, F. & Yasumura, M. 2000 A Speech Synthesis System with emotion for assisting communication. In *Proceedings of the ISCA Workshop on Speech and Emotion*, 167-172

[6] 徳久良子，乾健太郎，徳久雅人，岡田直之．言語コーパスにおける感情生起要因と感情クラスの注釈づけ 2001 人工知能学会研究会資料 SIG-SLUD-A003-2 9-14

[7] Russell, J. A., Lanius, U. F. 1984 Adaptation Level and The Affective Appraisal of Enviroments. *Journal of Enviromental Psychology*, 4, 119-135

[8] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J.G. 2001 Emotion Recognition in Human-Computer Interaction. *IEEE Signal Processing Magazine*, Vol.18, No.1, ISSN 1053-5888 32-80

[9] Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M & Schröder,M. 2000 'FEELTRACE': An Instrument For Recording Perceived Emotion In *Real Time. In Proceedings of the ISCA Workshop on Speech and Emotion.* 19-24

[10] 菊谷麻美，小川時洋，鈴木直人 1998 感情語の 2 次元空間内の布置について．同志社心理，No.45, 31-37

[11] 宇津木成介 1993 音声による情動表出と非言語的な弁別手がかり 異常行動研究会 (編) ノンバーバル行動の実験的研究 川島書店 Pp.201-217

[12] 坊農真弓 2001 音声対話における感動詞・応答詞の感情的意味機能—「ああ」を手がかりに 社会言語科学会研究大会予稿集 113-118