

対話における「気持ち」の定量化手法の開発

代表研究者	入野 俊夫	和歌山大学 システム工学部 教授
共同研究者	古山宣洋	早稲田大学 人間科学学術院 教授
共同研究者	花田里欧子	東京女子大学 現代教養学部 准教授
共同研究者	井上 雅史	山形大学 大学院理工学研究科 助教

1 研究の背景・目的

電気通信において重要になるヒューマンマシンインタフェースの一形態として、対話エージェントの研究開発がさかんに行なわれている。音声認識、音声合成、言語処理の基幹技術を中心として、Apple の Siri や NTT ドコモのしゃべってコンシェルに代表されるように実用化されている。しかし、それを使うユーザーには人間との対話とは全く異なった対応が求められる。また、言葉の内容以外に漂わせている、ユーザーの「気持ち」や「空気」を読んでもらうことを全く期待できない。それだけ難しい問題であることは明らかで、言語処理を中心としたアドホックなシステム開発を通じた試行錯誤をするだけで解決する見込みは高くない。そこで、言語以外の音響的／物理的な特徴量を用いて、相手の感情状態（「気持ち」）をある程度予測する指標の開発が目指されている。実際、国際会議において性能評価を行うためのコンペティションも開催されている[1]。また、「気持ち」の種類にかかわらず「気持ちを読めない」程度も定量化する指標を作ることができれば、現状ではごちない対話エージェントを、「気の利いた」ものにするヒントになり、今後の開発に役立つ可能性がある。

また、人間どうしの面接/対話においても、このような指標が役に立つ場面が想定できる。日常において、口振りや身振りを見逃して他人の「気持ち」を勘違いすることを、しばしば経験する。何気ない会話では問題とならないことでも、たとえば臨床心理の面談では、相談者（クライアント）の発している重要なサインを見逃すことにつながりかねない。また、逆に、クライアントに対する口振りや身振りが適切でないために誤解を与えてしまう場合も考えられる。そのために、臨床心理士養成課程においては、模擬面接等を専門家である教員と繰り返すことによりスキルアップを図っている。ここで、対話をビデオ撮影し、後で見返しながら内省することを通して行うことが効果的な取り組みの一つとなっている。しかし、教員からのアドバイスは、あくまで断片的な言語によるものであり、自らも気づきにくく言語化しにくい口振りや身振りを具体的な動きとしてフィードバックできるわけではない。そこで、上記のような口振りや身振りを通じた「気持ち」の定量指標値を対話の時系列に即して提供することができれば、教育効果も高くなると考えられる。

1-1 研究経緯

上記の背景の元に、当該研究グループは、数年前から対話におけるパラ言語情報/非言語情報と感情状態の関連性の研究を進めてきた。そこでは、京都教育大の臨床心理士養成課程のカウンセリング学習会において、対話の音声データ・身体動作センサデータ・ビデオデータの同時記録を継続的に行なってきた。さらに、和歌山大学でアニメーション説明課題の対話（深刻な内容としない目的指向対話）を収録し、計測システム系の開発や改良、データ分析法の開発を進めてきた[2]。また、対話のビデオを見ながら感情状態を連続的に評価して動的に入力する GUI (EMO system) を開発した。その上で、対話者本人および第三者 による動的な感情評価値を収集した。ここで扱う感情状態として、音声研究[3]で検討されてきた6つの評価項目（「快-不快」、「覚醒-睡眠」、「支配-服従」、「信頼-不信」、「感心-無関心」、「肯定的-否定的」）を対象とした。また、分析に必要な前処理（例えばテキスト書き起こし）も含めてデータベース化して蓄積した。さらに、音声・センサデータに様々な信号処理手法を適用して物理特徴量（たとえば基本周波数変化やうなずき度合）を抽出し、話者間や特徴量間の 相互関係の分析を実施した。この結果：(1) 対話者間の感情変化を誰もが同じように評価する「共有感覚」があることを客観的に示した。(2) 対話者本人が対話相手の感情を評価することを通して、自分の思いとの食い違いを見つける手段として利用できることがわかった。(3) 物理特徴量と感情状態の対応づけはまだあまり良くないものの、分析手法の改善の方向性がある程度明確 になった。この研究成果を元にして、本研究ではデータ収集、感情評価手法の改良と、データ分析手法の開発を行った。

1-2 本研究の目的

研究経緯をふまえ、以下の3項目に関する研究推進の提案を行なった。どれも相互に関連しているので期

間全体で行うこととした。具体的な内容と成果に関しては、次節で述べる。

- (1) データ収集とデータベースの構築
- (2) 感情評価値(目的変数)入力手法の改良
- (3) 物理特徴量(説明変数)からの定量化手法の開発

2 研究成果

2.1 データ収集とデータベースの構築

(a) 臨床心理面接のデータ整理とデータベース構築

研究分担者が、臨床心理士養成課程(京都教育大学)で定期的に主催していた【カウンセリング学習会】における臨床心理面接を対象とし、データ収集とデータベース構築を行なうことを進めることを目標とした。ここでの面接は、事前にストーリーが準備されている模擬面接ではなく、本当の悩みを相談する場である。なるべく人為的な要素を排除するためと、学習者にとって「相談者の立場になること」の機会から学習を進めてもらうためである。

本研究においては、助成期間前の2013年度までに収録した臨床心理面接のデータに関し、整理を行うことから開始した。理由の一つは、その時点で44件(対話)の面接データのほとんどが未整理であったため、新たな収録よりもデータ整理が優先と判断したためである。もう一つの理由として、臨床心理面接を実施している研究分担者大学を移籍したため、倫理委員会の手続き等がさらに必要で、学習会を立ち上げて落ち着いて収録できる状況ではなかったこともある。

初年度(2014年度)は、データに付与するラベル情報について、何が最も明確にされなければいけないかという観点から、専門分野の異なる共同研究者4名でブレインストーミングを何度か繰り返して議論を行なった。このラベルこそ、面接の状況や「気持ち」を定量的に評価するための重要な指標である。それと同時に、機械学習等で予測できる指標とするためには、比較的再現性良く安定に取れることが不可欠となる。人間による評価値のばらつきが大きければ、何を予測しようとしているのかわからなくなる。議論の結果、重要な観点として「傾聴」の概念(相手の話を深く聴いているか)の抽出を一つの目標とすることで合意できた。この方針を受けて、1つの対話に関して、2.2節で述べるEMObar(改良した感情推移観測システム)を用いて、面接のビデオを見返してセラピスト/来談者のそれぞれが相手の話を「聴いている/聴いていない」という連続評価量を入力することを、共同研究者4名がそれぞれ試みた。この結果、すべての評価者間でかなり大きな違いが出てくることがわかった。単に、臨床心理専門家と非臨床家による評価との違いという程度の相違ではなかった。全員の評価が終わった後の内観報告で、どのような点に基づいて評価を行なったかをそれぞれ述べた。この結果、「聴いている」という言葉の解釈の多義性が判明した。文字通り「聴いている」という意味(非臨床家)から、「相手の立場で言わんとしていることを解釈しつつ、さらに引き出そうとしている」という意味(その面接を行なった臨床家)までの違いがあった。また、おそらく臨床家の中でも、面接を担当したかどうか、個人の志向、所属する臨床心理の流派によって解釈が異なると考えられた。この結果自体は興味深いが、このデータを目的変数として設定して、予測することは困難である。

この結果を受けて、臨床心理の流派にかかわらず、慣れれば誰でも同じような付与できる、ある意味表層的なラベルをまずは導入することを考えた。ここで着目したのは、マイクロカウンセリング手法[4]で導入されている評定ラベル(たとえば、かかわり行動の有無、質問、発話の促し等)である。試行的なラベル付けを行なった結果、誰でも比較的容易にできることがわかった。そこで、発話内容の書き起こしおよび、このマイクロカウンセリングラベル付与を、臨床心理養成課程の学生に研究補助として行なってもらった。これに関しては、クロスチェックも含め初期に目標とした6対話に関して完了した。その後、さらに詳細なチェックを重ねた所、発話ごとにつけられたラベルは、評価者ごとにそれほど大きな違いがないことがわかった。しかしながら、同時に、ラベル自体よりも発話の区切りをどこにするかが評価者ごとに異なり、かなり難しいことがわかった。1発話ごとにラベルを付けるため、発話区切りの違いによる差異が生じてしまう。このゆらぎが生じてしまう問題を解消するために、雑誌編集等も行なった経験のある一名に作業を依頼することとなった。また、研究グループメンバーといっしょに、1対話分のビデオを通して見ながら書き起こし作業を行うワークショップ形式で、意図の共有化を図った。ここで、音声映像アノテーションツールELAN[5]に、発話書き起こしとともに、ラベルも載せるようにした。さらに、2016年に入って臨床心理の博士後期課程の学生もこの作業に参加してくれるようになり、順調に進むようになった。この結果、上記6対話のゆらぎの

あった結果の補正作業が終了した。定期的に入力作業を依頼できる状況になったため、今後1ヶ月に3対話程度はできると見込んでいる。さらに、EMObar (2.2節参照)を用いて、2~3名の臨床家による「傾聴」の度合いの連続評価を行なっていく予定にしている。本研究期間を通して、様々な問題点が生じ、解決のための議論と対策を重ねた結果、データ整備がスムーズに行なえる状況になったと考えられる。

(b) データ収録機器/システムの改善

助成期間前の2013年度までに、44件の臨床心理面接データを収集していた。2013年度においては、ビデオカメラ1台、6ch録音機器、頭部運動計測用の3軸加速度MEMSセンサと口元マイクを付けた超軽量ヘッドセット、ラップトップPCをセットにして使用していた。データの収録は順調に進み利用出来るデータを集めることができた。

この収集を通して、何点かの改善点が明確になった。まず、機材が複数あり、操作の順序を若干間違えたり、ラップトップPCの設定にまつわる小トラブルが発生していた。また、収録操作は開発を担当した研究代表者が主に行なったが、将来的には地元の医院等の臨床現場等でも、誰でも気軽に使えるようすることを目指している。そのためには、できるだけ簡便なシステム構成や操作法にすることが必要である。また、3軸加速度MEMSセンサを用いていたため、首を縦に振る動作はデータとして取れるが、首を横に振る動作(加速度があまり変化しない頭部横回転)を取ることは難しかった。さらに専用基板を用いたハードウェアを使っていたため、拡張性に乏しかった。特に、現状の2チャンネルを超えるチャンネル数にして、頭部だけではなく体の動きを同時収録することや、複数人での対話における各人の動きを収録することはできなかった。収集情報が多いほど説明可能性は高まる。この状況を受けて、将来に向けての拡張性確保のため、システムを要素技術から見直した。

まず、頭部運動収録系を、9軸(3軸加速度+3軸ジャイロ+3軸コンパス)センサ(InvenSense社MPU-9250)に変更した。これはI2C方式で制御できる。その制御用マイコンにはプログラミング環境の整ったArduinoを使うことにした。このうちUSB接続が可能で最も小型と考えられた、互換機のPololu社32U4A-star Microを採用した。また、センサとマイコンで電源電圧が異なるため、I2Cレベル変換回路を間にはさんだ。最終的に超小型(カバーを含め約25mm x 15mm x 12mm,数グラム)のセンサ系となった(図1)。これを細いUSBケーブル(2.5mmφ)でコントロールPCに接続できるようにした。USB micro-Bプラグの大きさもあって、従来用いていたセンサ系よりも若干大きくなった。しかし、ヘッドセットとしていたカチューシャに取り付けて装着しても、違和感が無いことを確認した。これにより、センサ用の専用外部ハードウェアが不要となったことは、拡張性を高める上で最大の利点である。なおワイヤレス化も検討したが、高品質音声録音のための小型口元マイク(DPA4060)が有線接続でしか扱えないこと、電池と送信回路の大きさの問題、安定性の観点から採用しなかった。

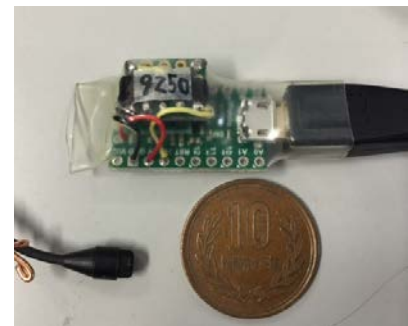


図1 9軸MEMSセンサと口元マイク

コントロールPC側のソフトウェアとして、MATLAB、PureData、Processingで実際にコードを書いて試すことを行った。MATLABは簡単に実現できたが、将来配布する場合にランタイムライブラリ等をインストールする必要がある。ソフトウェアの非専門家には敷居が高く、誰でも簡便に使えるという目的に合致しない。PureDataは、実時間処理でセンサデータを表示までできたが、パッチ形式のソフトウェア環境に慣れていないのもあり開発に時間がかかり、今後に不安が残った。Processingでは、収録に不可欠な時刻表示をPCの画面上に、データサンプリングと同期して行うことができた(図2)。また、フリーソフトで様々なOSに対応していることも含め簡便に使えると考え、最終的に採用した。さらに、センサ数を自由に変更できるようにプログラムを作り、実際に4つのセンサで同時収録が安定にできることと、60fpsでも時刻を正確に表示できることを確認した。非専門家でも簡便に利用できるように、センサをUSB接続すると自動的

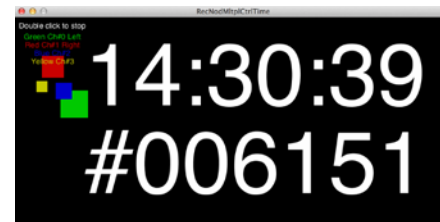


図2 センサデータ収集制御と時刻表示ソフト(左側の4つの四角形が4つのセンサ情報をフィードバック。下の6桁の数字は、フレーム番号。)

に数を検知する等の工夫も様々行なった。このシステムの回路図、ソフト、実装方法等の knowhow も必要に応じ公開できるので、他研究プロジェクトでも容易に活用できるものとする。

(c) 言語聴覚士養成課程における模擬面接対話の収録

上記のセンサも含めた収録機材を最適化の観点から見直し、実際に対話収録も行った。2016年2月に県立広島大コミュニケーション障害学科において、言語聴覚士養成課程の学生の演習で行なっている、難聴者の初回面接（インテーク面接）の模擬会話の収録を行なった。これは、難聴者にとって明瞭な音声の発話をどのようにすれば良いかを調べる研究である。これも目的指向対話であるため、直接的に「気持ち」を扱う課題ではないが、「明瞭性」の連続評価も行なうことも視野に入れているため、今回のプロジェクトにも関連する。また、収録システムの改善のためにも役に立つ。

この模擬面接では、言語聴覚士1名、高齢患者役1名、付き添い男性役1名の計3名の対話を収録することになった。そこで、上記9軸MEMSセンサ系を3セット用意し、音声に関しても3人分収録することになった。ここで用いた、収録システムの構成は次のようになった。(1)ビデオカメラ(4chステレオ録音可)Zoom Q8, (2)複数音声録音Roland Octa capture, (3)顔正面映像収録 Ricoh Theta S, (4)パソコンMacbook Air, (5)口元マイクDPA 4060, (6)上記9軸MEMSセンサ。これらのセットアップや練習セッションを何度か行なった結果、本番では成功裏に収録を行なうことができた。現在、この音声の分析をopenSmile[1]で開始しようとしている段階である。本研究によって、今後の収録に活用できる knowhow を得られた。

2.2 感情評価値(目的変数)入力手法の改良

本研究期間以前に、感情状態を連続的に2次元座標値としてマウス/パッドを使ってGUIで入力するEMO system (感情推移観測システム)を開発し利用してきた[2]。これは、音楽の印象/感情評価用として開発されたGUIであるEMuJoy[6]を発展させたものである。これを用いて、対話者それぞれの感情評価値を入力していた[2]。ところが、本研究における評価の結果、問題点も見いだされた。そこで、改良したユーザーインタフェースを開発し、その評価を行なった。

(a) EMO システムによる評価再現性

我々の行なった先行研究[2]と、まったく同じ2次元座標値マウス入力のEMOsystem(以下ではEMO2Dと呼ぶ)を用いて、同じデータで同一被験者が複数回入力する場合に、感情評価値に再現性があるかを確認する評価実験を行った[8]。この場合の評価軸には3つの組み合わせを用いた:(1)x軸「快-不快」-y軸「覚醒-睡眠」、(2)x軸「支配-服従」-y軸「信頼-不信」、(3)x軸「感心-無関心」-y軸「肯定的-否定的」。さらに、心理尺度構成法の観点から、評価入力軸に度合いの形容詞ラベル(かなり、やや等)を付与した版も作成し、比較も行った。再現性がある程度担保されることと、ラベルにより天井効果が無くなることはわかった。ただし、2次元のx軸、y軸の間の相関が高いように見えることは、以前の状況と変わらなかった。しかし、2軸の間に必ずしも高い相関があるとは考えられない。特にEMO2Dの元となったEMuJoy[6]においてはx軸「快-不快」-y軸「覚醒-睡眠」上の平面で定義されるRussellの円環モデル[7]の範囲のみで評価していた。これに対し、EMO2Dでは軸の定義も評価に合わせて変更した。すなわち、軸は定義できるものの、その間を2次元平面と見なして扱うことができない場合に適用してしまった可能性がある。そこで、入力手法を変更した新規のEMO systemを開発し、評価することを行なった。

(b) 並列型感情推移観測入力装置 EMObar の開発

上記の理由ばかりではなく、2次元入力では、横方向と縦方向の値自体を意識して操作することが難しい。被験者は、おおまかな平面上の位置をだいたい良さそうと思って入力していると考えられる。軸で表現される値を意識させるためには、むしろ2つの評価の独立で行えた方が良く考えた。そこで、スライド抵抗器を左右に平行に並べてつまみの位置を評価値として出力し、画面にも直感的にバーの大きさで表示するユーザーインタフェース(UI)、EMObarを開発した(図3)[8]。入力装置部分は、スライド抵抗器2本を並列に並べたシールド基板をArduino

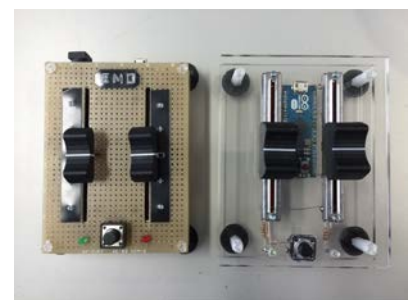


図3 新規開発の感情推移観測システムの入力装置 EMObar
(左)初期版, (右)最新版

に載せた簡単なものである。これをパソコンにUSB接続し、パソコン側でデータ収集する。このプログラム言語にはProcessingを用い、ビデオ映像用の画像フレームにオーバーラップされて、入力値を反映したバーグラフが表示されるようにした。被験者は、スライド抵抗のツマミの位置が反映されるバーの長さを参照しつつ、時々刻々感情評定値を入力できるようになっている。この並列型の入力装置であれば、左右の手で独立に操作できるため、評定軸間で本来無いはずの相関が出ることは無いと期待できる。

(c) 数字追従入力タスクによるインターフェース評価

EMObarと従来法であるEMO2Dとのインターフェースとしての正確性を判断する実験を行なった[8]。画面に3秒ごとに表示される数字に追従して、値を入力する実験タスクである。このタスクとしたのは、感情評価であると個人内・個人間でのばらつきが大きく、インターフェースの評価としては適切とはないと考えたためである。逆に数字追従入力適切に行なえば、感情の内観値を的確に反映した入力ができると考えられる。

図4左に、実験におけるEMObarの操作状況、図4右に表示画面の様子を示す。図5左には、EMObarの場合の数字表示と、スライド抵抗値を反映したバーグラフ表示を示す。左の緑色の数字に対応させて左側のスライド抵抗を操作し、左側の緑色のバーグラフを目的とした値まで持っていく。右の赤色の数字についても同様である。3秒ごとにこれらの数字が±1以内でランダムに変わっていく。この数字になるべく早く追従し、バーグラフの値を目的値にするタスクとなっている。図5右はEMO2Dで、同様に数字に対応した(x, y)の値になる点にポインターを持っていくタスクとなっている。数字の大きさや色は同じにした。ターゲット値と入力値の差が0.5以下で正解とした。

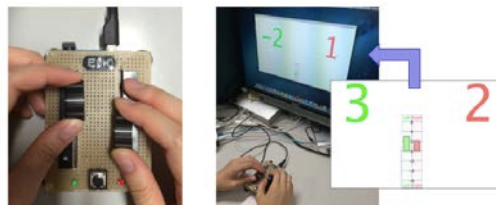


図4 EMObar 操作と数字追従入力タスク



図5 数字追従入力タスクにおける数字表示
左図：EMObar の入力バーグラフ表示
右図：EMO2D の2次元入力座標表示

セッションは、練習1(30秒間)、練習2(60秒間)、本実験(3分半)2回で構成した。練習1、練習2でそれぞれ100%正解した段階で、本実験に進み正解率や誤差を計算した。2種類のユーザーインターフェースを使う順番は、グループを均等に分けてバランスさせた。実験は21歳から24歳の大学生、大学院生の男女計12名であった。実験の目的等を書類と口頭で説明し、同意書を得た。

まず、練習1,2で正解率100%になるまでの回数の平均と標準偏差は、表1のようになった。明らかにEMObarの方が、少ない練習回数で完全正当ができるようになったことがわかる。EMO2Dで分散が大きいのは、15回あるいは26回の練習でやっと完全正当できた被験者も居たためである。しかし、それを除いたとしても平均的に高かった。確認のため、全データで入力手段、被験者、練習番号に関する3因子分散分析を行った。入力手段の主要因(F(1, 34)=14.13, p<0.001)、被験者の主要因(F(11, 34)=2.42, p<0.05)に有意差があった。練習番号の主要因(F(1, 34)=2.21, p=0.15)には有意差は無かった。以上のことより、EMObarの操作の方が容易に修得できることがわかった。

表1 正解率100%になるまでの練習回数の平均(標準偏差)

	練習1	練習2
EMO2D	4.0 (2.2)	6.7 (7.1)
EMObar	1.7 (1.0)	1.8 (1.0)

本実験では、ほとんどの場合正解率が100%であることを想定していた。しかし、EMO2Dでは必ずしも正解率が100%とはならない場合も数多くあった。そこで目標値と入力値の差分について分析した。この結果、EMObarでは平均値が0.2以上ずれることはなく、標準偏差もほとんどの場合0.2未満で収まっていた。これに対し、EMO2Dでは、平均で0.2以上、標準偏差が1以上の場合も数多くあった。この結果は、上記で練習回数が多かった被験者だけではなく少ない被験者でもそのような場合に該当することがあった。EMO2Dの場合、一度入力がずれてしまうと修正がむずかしい問題があり、これが反映したものと考えられる。

以上の結果から、数字を追従入力するタスクにおいては、新規開発した並列型の入力装置EMObarの方が有利であると結論づけることができる。感情入力の場合も、感情の度合いを内観の値として持っていれば、EMObarの方が忠実に十分早く入力することができるものと考えられる。そこで今後の実験では、必ずしもRussellの円環モデルの2軸(「快-不快」と「覚醒-睡眠」)を使うわけではないこともあり、EMObarを使う

ことにした (2.1(a)節)。

(d) EMObar の音楽印象タスクによる評価

上記の検討では、数字追従タスクで並列入力の EMObar が 2 次元入力の EM02D よりも有利であることを示した。しかしながら、実際の感情入力において、従来研究と同様に利用できることを直接示したわけではない。そこで、原点にもどって、EMuJoy が利用されていた音楽の印象評価に関して、EMObar で同じ評価値を出せるのかの検証実験を行った[9]。本研究では対話を対象としているが、その感情評定値の連続評価に関して、第三者による客観的なデータの蓄積は無い。これに対して音楽の印象評定に関しては、EMuJoy によるデータが論文にも掲載されていて直接比較できる。評価対象は異なっているが、入力手段の評価のためにはこちらの方が確実である。また、EM02D は EMuJoy を元に作った 2 次元入力手法なので、今回の比較では用いる必要はないと判断した。なお、上記の数字追従タスクでは、逆に EMuJoy では数字を画面上に同時表示できなかったため対応できなかった。

使用する曲は、EMuJoy を使用した先行研究で時系列データが残っている 6 曲を選んだ [10, 11, 12]。曲の長さは 3 分前後とした。3 分を超える曲は、先行研究のデータから特に感情の動きが大きいと思われる部分を抽出した。本実験の被験者は実験を行うのに問題のない聴力を有した 21 歳から 22 歳の大学生の男性 7 名、女性 3 名の計 10 名であった。説明書と口頭で実験の目的や実験を行い、同意書を得た。

評価尺度には、先行研究と同じ、「快-不快軸」と「覚醒-沈静軸」を用いた（上記の「覚醒-睡眠」は直訳の和訳で、ニュアンスが伝わりにくかったため変更し「覚醒-沈静」とした。）また今回の実験では、EMuJoy, EMObar による 2 軸同時評価、EMObar による 1 軸ずつ分けて評価を行なう場合の 3 種類の評価手法を行なった。本番の実験に先立ち、評価尺度を説明し、練習のための曲 (67 秒) を用いてユーザーインターフェースの操作に慣れてもらった。疑問点がないことを確認した上で本実験を行なった。曲やインターフェースの順番は 10 名でバランスが取れるように配置した。

EMuJoy を用いて感情評価実験を行った先行研究[6]の時系列データと、今回の実験で得られた時系列データを比較する。実験の結果から得られた時系列データを比較する。図曲名“Chocolat”について、先行研究[6]に報告されている EMujoy で評価した全被験者の中央値(図 6 左)と今回の実験で得られた値の中央値(図 6 右)を示す。軸のスケールが異なるが、概ね概形が似た結果が得られていることがわかる。また、今回の実験の 3 条件間でも類似の結果が得られていることもわかる。先行研究[6]のデータは図でしか提供されていないため、数値的な相関はまだ取っていないが、高い相関が得られるものと考えられる。他の曲や評価軸でも概ね曲線は類似している。今後、図から値をデジタル化して、定量分析することを予定している。

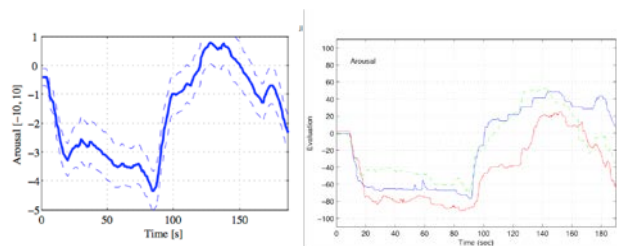


図 6 曲名“Chocolat”の覚醒-沈静軸での評価結果。左図) Nagel らが報告した結果[6]。右図) 今回の実験で得られた評価値。赤) EMuJoy, 青) EMObar による 2 軸同時評価、緑) EMObar による 1 軸ずつ評価。

以上のことから、EMObar でも EMuJoy による評価と相関の高い結果が得られることがわかった。このことにより、EMObar でも従来法と同じような感情評価入力ができることを示せたと考えている。さらに、Russell の円環モデルに対応する感情（「快-不快」と「覚醒-沈静」）を入力する場合においても、2 次元入力が必須というわけではないことも示唆している。

2.3 物理特徴量(説明変数)からの定量化手法の開発

第 3 の目的としていた、物理特徴量(説明変数)からの定量化手法に関しては、初期検討段階で研究期間終了を迎えた。音から openSmile[1]を使つての特徴量抽出は開始し、目的変数へのマッピングは線形回帰とステップワイズ近似を用いて予備的に検討した。実際にアノテーションが付与されたデータが多量にあれば、機械学習(マシンラーニング)等の最近の統計手法で行なうことを考えていた。ところが、前節までで述べたように、データ整理に関して当初想定した以上の問題が生じて対応を行なったため、最終的に統計分析に乗せられるほどの十分なアノテーション量が確保できなかった。また、感情評価値(目的変数)入力手法の改良や評価も、目的変数を的確に決めるためにも重要な過程であるため、かなりの時間を使った。これらの

前段階の盤石な基盤が無い限り、いくら分析しても意味が無い。上記 2.1、2.2 節における検討はしっかりでき、元来難しい問題の一部を解きほぐすことはできたと考えている。この知見に基づき、順次定量分析に掛けられるデータは増える予定である。今後、データがある程度揃った段階で、一気にシミュレーションを開始する予定である。

3 まとめ

本研究においては、感情/感性の客観的推定に迫るため、(1) データ収集とデータベースの構築、(2) 感情評価値(目的変数)入力手法の改良、(3) 物理特徴量(説明変数)からの定量化手法の開発に関して提案を行い、研究を進めた。(3)は検討段階までしか進めなかったが、その基盤となる(1)(2)は進めることができた。(1)は、方針策定に時間がかかったが、今後順調に進めるだけの方向づけができた。また、データ収集手段に関しては格段に進歩させることができた。(2)のインタフェース開発は、評価も含め十分な進捗と成果があったと考える。これについては学会発表と論文化をする予定である。今後、当該研究課題の成果を元に、感情/感性情報である「気持ち」を定量的に推定する研究をさらに進める所存である。

【参考文献】

- [1] Schuller, B.W., "The computational paralinguistics challenge," IEEE Signal Processing Magazine, pp. 97--101, Jul., 2012.
- [2] 八木みゆき, 森田礼子, 中井正人, 西村竜一, 河原英紀, 入野俊夫, "説明対話における韻律情報および頭部運動と感情評価値の関連性分析の試み," 電子情報通信学会/音響学会 音声研究会, SP2013-67, pp. 15-20, 千葉大, 千葉, 2013年9月18日~19日, (発表日 9月18日) * ±
- [3] 森大毅, 相澤宏, 粕谷英樹, "対話音声のパラ言語情報ラベリングの安定性," 日本音響学会誌, 61(12), pp. 690-696, 2005.
- [4] 福原真知子, アレン・E. アイビイ, メアリ・B. アイビイ, "マイクロカウンセリングの理論と実践—事例場面から学ぶ," 風間書房, 2007.
- [5] 細馬宏通, "ELAN即席入門," http://12kai.com/elan/elan_abc.html (最終閲覧日: 2016/5/29)
- [6] Nagel, F., Kopiez, R., Grewe, O. & Altenmüller, E. (2007). 'EMuJoy' - Software for continuous measurement of perceived emotions in music: basic aspects of data recording and interface features. *Behavior Research Methods*, 39(2), 283-290.
- [7] Russell, J.A. "A Circumplex Model of Affect", *Journal of Personality and Social Psychology* 1980, Vol. 39, No. 6, 1161-1178.
- [8] 柴田知紗, "対話における感情評価値を連続的に入力するインタフェースの改良", 和歌山大学システム工学部卒業論文, 2014年度.
- [9] 坪田遥香, "音楽聴取時の感情の連続評価におけるインタフェースの比較", 和歌山大学システム工学部卒業論文, 2015年度.
- [10] Nagel F, Kopiez R, Grewe O and Altenmuller E (2009), " "Worms" in (E) motion: Visualizing Emotions Evoked by Music ", In Audio Engineering Society Convention 127.
- [11] Grewe O, Nagel F, Kopiez R and Altenmuller E (2007), "Emotions over time: synchronicity and development of subjective, physiological, and facial affective reactions to music.", *Emotion*. Vol. 7(4), pp. 774. American Psychological Association.
- [12] Grewe O, Nagel F, Altenmuller E and Kopiez R (2009), " Individual emotional reactions towards music: Evolutionary-based universals? ", *Musicae Scientiae*. Vol. 13(2 suppl), pp. 261-287. SAGE Publications.

〈発表資料〉

題名	掲載誌・学会名等	発表年月
Continuous annotations for dialogue status and their change points, (Masashi Inoue, Toshio Irino, Ryoko Hanada, Nobuhiro Furuyama, and Hiroyasu Massaki)	10th workshop on Multimodal Corpora: Combining applied and basic research targets, LREC 2014 workshop.	2014年5月27日
Multimodal dialogue segmentation with gesture post-processing (Masashi Inoue, Kodai Takahashi)	The 9th edition of the Language Resources and Evaluation Conference (LREC), 26-31 May 2014	2014年5月29日
The Development of Solution Building Inventory Japanese version: Validation of the SBI-J3 (Takagi, G., Wakashima, K., Sato, K., Ikuta, M., Hanada, R., & Smock, S. J.)	International Journal of Brief Therapy and Family Science, 5(1), pp.19-35	2015年
CSIの信頼性と妥当性の検討—大学生を対象として— (小林大介, 高木源, 清水優, 安藤樹, 赤間由依, 若島孔文, 安保英勇, 花田里欧子, 生田倫子, & 佐藤宏平)	日本ブリーフセラピー協会	2015年9月6日
Interactional Mind 8 (2015) ブリーフセラピーテキスト&ワーク (日本ブリーフセラピー協会 (編) 花田里欧子他著)	東京: 北樹出版	2016年