コトダマ：感情認識を用いた音声入力による インタラクティブな照明の制作

岡戸雄一郎[†1]　串山久美子[†2]　栗原渉[†2]

概要：人の発声する同じ言葉でも喜怒哀楽などの情動感情を伴うことで意味の違いを生じることがある。また最近ではコロナウイルスの流行により、人とコミュニケーションをとることが極端に少なくなり,感情を表に出さずに過ごす日々が多くなった.筆者らは人の発音の情動情報を日常的に使用するモノへの視覚化に注目し,人と接することなくとも感情を表に出せる状況を作れないかと考えた.本研究では音声認識と感情解析を使用しユーザーの肉声から感情値を読み取り,感情を反映させた色に光るインタラクティブな照明を制作した．人はモノに対して通常は感情を伴わず言葉を発するが,音声認識に情動要因を付加する事でユーザーの発声体験や情動デバイスとしての照明に自分の感情を意識した体験価値が産まれると考える.

1. 導入

人間は生き物の中でも感情を表現するのに長けている. 感情を表に出すということは良い人間関係を構築するにあたって必要不可欠である. さらには心のバランスをコントロールすることにもつながる．例を出すと楽しい時は笑う, 悲しい時は泣くといったように感情を出すだけで心のバランスが取れ精神を安定させられる．しかしコロナウイルスの流行で今年は家で自粛生活がすることが大半であり, 人とコミュニケーションをとることが極端に少なくなった. つまりは感情を表に出さずに過ごす日々が多かったのではないだろうか． 筆者らは, 人と接する事なくとも 1 日の中で自然と声を出し感情を表に出せる状況を作れることはできないだろうかと考える.

また声は人間にとって最も自然なコミュニケーション手段である. したがって, 人間と機械のインタフェースとして音声を使いたいという要求は強い. このような背景のもと, 機械による自動音声認識の技術は発展してきた。最近では, Amazon Echo（Alexa）や Google アシスタント、Apple の Siri などのスマートスピーカーの存在も日常的に使われるようになり、AI による音声認識が進化したことで、声だけで機械を操作したり、会議の議事録を効率よく作成するなどの実用性が表立っている。人の口から発する言葉には感情がこもっている. 同じ言葉でも話し手の感情や言い方で相手への伝わり方や意味までも変わってくる.

しかし現代の音声認識を用いた AI は音声をデジタル化し, 単語の並びと理解する. そこにユーザーの感情に対する一切の考慮はなされておらず, どんな感情で指示をしようが AI は言葉の条件を満たせば指示を実行する. 本研究では人間の肉声に込めらた感情に着目し, それを入力媒体とする照明を制作した.

2. 先行研究

“Hidden Words of Noise and Voice”

本作品は観客が HMD を装着し，入力された音声から声の音高,ボリューム,ピッチ を抽出して，それに応じた抽象的な形態が仮想空間中に生成され,動きまわる.音声入力された肉声は単語や文脈ではなく,声の音高,ボリューム,ピッチといったところに焦点を当て独特な世界観で可視化しているのがこの作品の魅力である.ユーザーの音声入力に対し, 実用性からは離れ新たな体験を提供すると共に声のパラメーターをインタラクティブに表現している点よりこの研究を本研究の先行研究とする.(図 1)

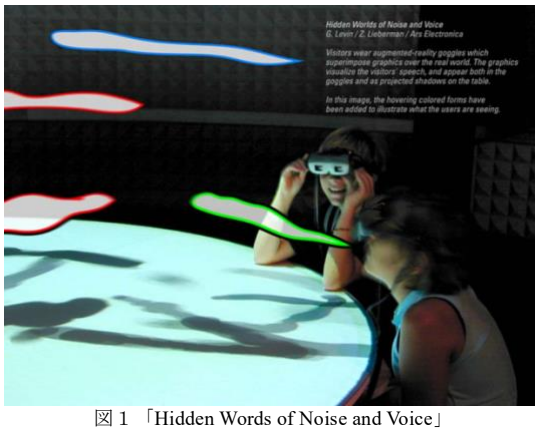

図 1 「Hidden Words of Noise and Voice」

3. 作品制作

3.1 音声認識と感情の考察

音声認識をするにあたり感情といった要因は本来解析するのに余計であり,AI は感情によってプログラムを左右されない.しかしもし仮にも感情があると仮定するとどうだろうか.人に無感情で指示された場合それを実行しようとは思わなかったり,怒って命令された場合は機械が腹をたててしまったりする.通常時は人は音声認識 AI に対して感情を伴わない必要最低限な言語情報でしか音声入力をしな

†1 東京都立大学大学院　システムデザインシステムデザイン専攻 インダストリアルアート学域
†2 東京都立大学 システムデザイン研究科

いが,感情を要因に組み込む事で今まで意識していなかった人と音声の関係が生まれるのではないだろうか..

3.2 システム

本研究では,音声入力された肉声から感情を分析するにあたって Empath という API を用いた. Empath は、音声等の物理的な特徴量から気分の状態を独自のアルゴリズムで判定するプログラムである。数万人の音声データベースを元に喜怒哀楽や気分の浮き沈みを判定する.

また Empath を使用するにあたり感情を解析するには音声入力された肉声は指定(図２)のフォーマット化したものでなければならない.したがってこれらの動作を一連で遂行しリアルタイムの音声認識を実現するため, Python でマイクから入力された肉声を API 指定のフォーマットに保存し,Empath で解析するシステムを構築する.それらの解析された感情の値を OSC 通信で MAC から ESP に無線でつなぎ RGB 数値に変換し LED を光らせる.照明は数ヶ所光らせる場所があるためテープ LED を用いた.

- ●PCM WAVE形式、16bitであること.
- ●データサイズが1.9MB以下であること.
- ●録音時間が5.0秒未満であること.
- ●サンプリング周波数が11025Hzであること.
- ●チャンネル数が 1(モノラル)であること.

図２　Empath で解析できる音声形式

3.3 感情の色彩化

ESP には予め Mac から送られてきた感情の値を受け取り RGB の数値に変換するプログラムを Arduino で書き込む.Empath で解析する感情は calm「穏やかさ」anger「怒り」joy「喜び」sorrow「悲しみ」energy「元気度」の５種類あり,それらの検出できる数値の範囲は０から５０である.これらの感情値を色彩化するにあたり, 心理学者ロバートプルチックのプルチックの感情の輪(図３)を参考にする.プルチックの感情の輪とは人間の感情を色相間のように表現したもので, 怒り、恐れ、期待、驚き、喜び、悲しみ、信頼、嫌悪の８つの基本感情からなる.感情の分類は色の分類と似ており,色も組み合わせれば無限に通りがあるように,感情も組み合わされば新たな感情を生み出す.

今回はこれを元に感情の値と RGB 値を紐付ける.RGB の値の上限は 255 であり,それらの値の求め方は計算式図４に示す.R の値は赤色と黄色に位置している怒りと喜びから anger と joy の値で求める.G の値は黄緑と黄色に位置してる喜びと信頼から calm と joy の値で求める.B の値は青色と赤色に位置している悲しみと怒りから sorrow と anger の値で求める.またプルチックの感情の輪において感情の強さは色の濃さに比例すると考えられる.これより本研究では energy の値を感情の強さに比例すると考え,照明の輝度と紐付ける.

図３　プルチックの感情の輪

R の値　(anger ＋ joy) ÷2 ×255/50
G の値 (calm ＋ joy) ÷2 ×255/50
B の値 (sorrow ＋ anger) ÷2 ×255/50

図４　感情の色彩化の計算式

3.4 実装

言霊というのは、発した言葉が音としてだけではなく力をもち、その言葉のきっかけで現実に何かしらの影響を与えるという物である.

本研究では、そんな言霊から発想を得て球体型照明の「コトダマ」を制作した.照明は 150mm の球体型である.下半球に 4 つの支えがついており,照明内は空洞でテープ LED に ESP と携帯電源機を接続して配置されている.照明の上半球は PLA 素材を薄さ 2mm で 3D プリンターで出力し,照明の色彩変化が視認しやすい薄さに調節した.

また下半球は木製素材で出力し,「コトダマ」の目に該当する直径 15mm の穴を２箇所設ける.目部分が光るようにテープ LED を内部に貫通している穴の前に配置する.目の部分には下半球と同様の木製素材を薄さ 0.5mm で出力したものをはめ込む.(図５)

図５　「コトダマ」3D データ

実際に下記の実装図(図６)通りマイクにユーザーが音声入力すると,その音声データを Python で指定フォーマットに保存しする.それを Empath API の解析にかけて,取得した感情値を OSC 通信で照明内の ESP にデータを送る.ESP では送られてきた感情値を RGB の数値に変換して LED を光らせる.

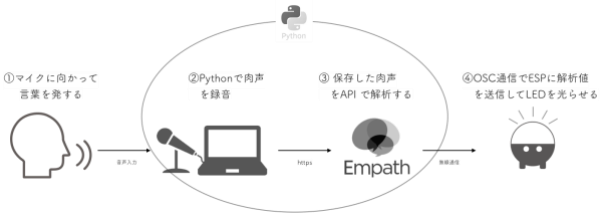

図６　システム実装図

3.5 ユーザーシナリオ

ユーザーはコトダマに向かってマイクを通して電気をつけるよう話しかける.その言葉にから検出された感情によって「コトダマ」のフィードバックは異なる.検出される感情は calm「穏やかさ」anger「怒り」joy「喜び」sorrow「悲しみ」energy「元気度」の５種類ありそれらの各値の幅は０から５０である.

実際に本研究の内容を把握していない男女 14 名で音声データを取得した.音声入力してもらう想定として①無感情②怒り以外の感情③怒りの感情の３つを想定して音声入力をしてもらった.それらの音声データを API で解析し,感情がこもっているか否か,また怒っているか否かの判断基準を設ける.①の場合の各感情の合計値の平均は２７.また②の場合の各感情の合計値の平均は５５.③の場合の怒りの感情値の平均値は３２であった.これより感情がこもっていないと判断する場合は各感情の合計値が①と②の場合の平均値の４１に満たなかった場合とし,対して感情がこもっていると判断する場合は,感情の合計値が４１以上の場合とする.また怒りの値が③の状況の３２を超えた場合は①②の条件より優先してユーザーが怒りの感情で話しかけていると判断する.これらを踏まえたユーザーの体験フロー(図７)が以下である.

感情がこもっていないと判断した場合,照明は点灯しない.感情がこもっていると判断した場合は本研究の感情の色彩化に準じた色(図８)の LED を光らせる.また怒りの感情で話しかけられていると判断した場合は赤色(255：０：０)に点滅(図９)し,まるで「コトダマ」が機嫌を損ねたかのような挙動を示す.照明は 10 秒間点灯する仕様であり,ユーザーが新たに音声入力をした場合は照明はその肉声の感情によって判断されたフィードバックを再度与える.

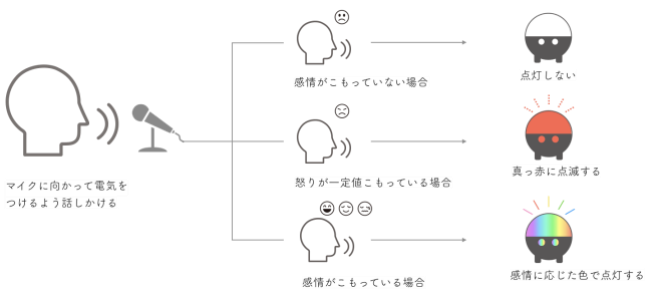

図７　ユーザー体験フロー

図８　感情がこもっている場合

図９　怒りの感情がこもっている場合

4. まとめ

4.1 考察

本作品は声で操作する既存のデバイスと違い,ユーザーの感情を読み取りそれによってフィードバックを変える新たな音声入力の体験を可能にするデバイスとなっている.無感情で軽率に光をつけるように話しかけるのならば照明はユーザーに対して一切反応をしない.そこから試行錯誤をし,感情を込めてデバイスに話しかける事で照明はそれに応じた色の光を灯してくれる.それらの一連の体験を通じて従来のユーザーと音声操作デバイスの関係ではなく,ユーザーはデバイスに対して,感情を伴って発言するようになる.感情を普段表に出す状況がない場合でも,この照明は感情を自然と表現できる機会を提供する事ができる.

また普段は意識しない自分の言葉にどのような感情がこもっているのか,また相手はどう感じているのかをこの照明を通した体験は再度考えるきっかけになるのではないだろうか.本研究で用いた API では所得できる感情値に制限があり,プルチックの感情の輪の基本感情である心配と驚きに対応する感情の抜けがあるため感情の検出面でも色彩化面においてもまだ不完全であると考える.

4.2 今後の展望

本研究では音声入力に制限があり,特に５秒未満でなければならないためユーザーの音声入力の融通が効かない.この問題に対して,５秒以上の録音に対しては５秒ごとに分けて WAVE 形式で保存し,それぞれを個別に API で解析するプログラムを組む事で長時間の音声入力も可能になると考えられる.

また音声感情認識においては人の声質によっての感情認識に多少のブレがある.特に男女によっての差異があるように感じた.深層学習を感情検出値のキャリブレーションをすることは今後必須であると考える.

本作品は感情を API で解析するため,PC にマイクを接続しており,マイクと照明とが別々になっている体験環境である.今後は照明にマイクとワンボードコンピュータを内蔵する事で体験環境が照明一つで完結しスマートになり,さらにプロダクトとしての実用性や魅力を引き出すことができると考える.

参考文献

1) 江渡浩一郎,アート・エンターテインメントにおける音インタフェース, 社団法人情報処理学会(2004)
2) 河原達也,音声認識技術の変遷と最先端,日本音響学会誌 74 巻 7 号(2018)
3) Empath Emotion AI in the age of the voice https://webempath.com
4) 宇津木成介, 基本的感情の数について(About the number of basic emotions) 神戸大学大学
5) Robert Plutchik: Emotion: Theory, Research, and Experience, New York: Academi
http://office.microsoft.com/ja-jp/products
6) Robert Plutchik, “The nature of emotions, ”American Scientist,Vol. 89, Iss. 4.