

Voicelot: 音声入力をエントリーとした 意思決定省力化システムの提案

眞野 隼輔† 渡邊 恵太†

複数人で食事の内容や目的地を決める際に、じゃんけんやくじ、抽選アプリなど偶然にまかせることで公平に決めることができる。一方、あらかじめルール決めや選出基準を決めておく必要があるうえに、人数が多いと意思決定に時間がかかってしまう。そこで本研究では、音声入力を用いた抽選システム Voicelot を提案する。参加者は発声によりエントリーし、自動で抽選することで意思決定を素早く円滑に行うシステムである。

1. はじめに

複数人の中から掃除の当番を選んだり、外食の行き先を決めたりすることに悩む場面がある。そのような場面で、じゃんけんやくじ、抽選アプリなどの偶然にまかせた手法を用いることで当番や行き先などを公平に決めることができる。しかし、これらの手法は事前の準備や合意形成を必要とし、人数が多ければ多いほど決定に掛かる時間が長くなってしまふ。例えば、 N 人が順位を決める、もしくは一位の人のみを決めるために必要なじゃんけんの回数の期待値は $(3/2)^N$ であり[1]、人数が多いほど決定するまでの時間がかかりがちである。また、後出しが起こりえる問題や勝った人から選ぶのか、順位で決めるのかなど、あらかじめ合意形成を必要とする。くじやサイコロを用いる場合でも、人数が多いほど準備や実行に時間がかかり、また同様に合意形成を必要とする。

一方、抽選アプリなどコンピュータを用いたものでは、決定自体に時間はかからないが、メンバーのエントリーに手間がかかる。金子らは、カメラと顔認識を用いて顔をエントリーとすることで、多人数であっても抽選のための事前準備を省力化できるシステム Facelot[2]を提案している。

本研究では Facelot と同様にこうした抽選のための事前準備を省力化した意思決定支援システム Voicelot を提案する。Voicelot は、Facelot と似た特徴を持ちつつも、参加者の声をエントリーに使用する点が特徴である。



図1 Voicelot の利用風景

2. Voicelot

Voicelot は音声入力を用いた抽選システムであり、参加者は声を発することで抽選にエントリーできる。Voicelot はウェブアプリケーションとして実装し、スマートフォンなどの端末に内蔵されているマイクを用いて利用することができる。エントリーは一人ずつ順番に、タイミングに合わせて発話を行うことで進める。発声のタイミングには山手線ゲーム（古今東西ゲーム）と同じシステムを採用した。二拍手と無音が2秒ずつ交互に再生される音声に合わせて、無音のタイミングで一人ずつ任意の発話を行う。具体例として、お昼ご飯に食べるものを決めたいときに1人目「うどん」、2人目「カレー」、3人目「寿司」、4人目「パスタ」とリズムに乗りながら発話していく流れを図2に示す。参加者全員のエントリーが終了した段階で、画面に表示されている抽選開始ボタンを押す。拍手以外の部分が個別に録音してあり、その中から一つがシステムによってランダムに選ばれ、再生されることで結果発表を行う。

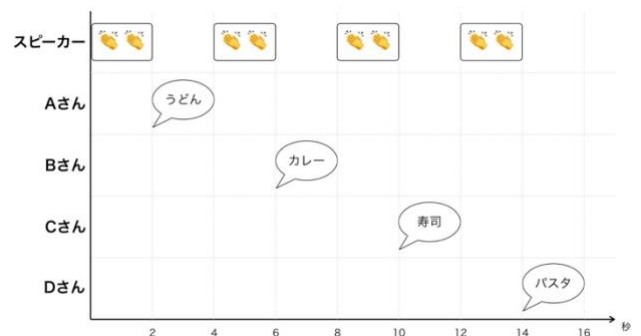


図2 Voicelot におけるエントリーの流れ

3. アプリケーション

Voicelot によるアプリケーション例を三つ紹介する。

3.1 録音によるランダム抽選

エントリーを拍手で区切ることで、参加者の発声を個別に収録することが可能となる。これらの中から、完全に無作為の一つの録音を選出し再生する。

† 明治大学
Meiji University

3.2 元気度抽選

元気度抽選は、声の物理的な特徴量を用いたアプリケーション例である。感情分析 API[‡]は音声データを分析し、発話者の元気度を 0~50 の範囲で予測値を返す。ある人の当たる確率を（その人の元気度÷全体の元気度の合計）として抽選を行うことで、元気よくエントリーした人ほど選ばれやすくなる。元気度抽選では、参加者自身が選ばれやすくなるために自発的に元気の良い声を出そうとすることから、抽選を盛り上げるエンターテイメント的側面を持ち合わせている。例えば、カラオケにて誰が最初に歌うかを決める際に元気度抽選を用いることで、参加者が自発的に元気な声を出してエントリーする。これにより、場の盛り上げや発声練習に繋がることが期待できる。

この他にも、感情分析 API では平常、怒り、喜び、悲しみを数値に変換することができるため、それぞれのエントリーに対して抽選確率の重み付けを行ったアプリケーションが考えられる。

3.3 音声認識を用いた抽選

音声認識 API[§]を使用することで、エントリーした人が何を発言したのかを文字に変換できる。何らかのテーマを元に抽選する場合、発言内容から機械が選ばれやすさを自動で算出して、それぞれのエントリーに対して重み付けを行う。例えば、昼食をどこのお店で食べるかを決める場面では、エントリー時に参加者が食べたいものを順に発言してもらう。発言内容から該当するお店が周囲に存在するか、現在地からの距離、お店の評価などを加味した上で抽選に重み付けを行う。また、抽選結果の発表時に、該当するお店を 1 店舗のみ提示することで、お店選びに掛かる時間を短縮することもできる。

4. 議論

4.1 利用シーン

Voicelot は声のみを使ったエントリーシステムである。そのため、料理中や自動車の運転中などの手を離せない場面であっても、ハンズフリーで抽選に参加できる。また、電話を用いた遠隔地からのエントリーも可能である。通話アプリケーション内の機能として実装すれば、会話の中で意思決定をスムーズに行うことができる。

抽選結果の発表にもエントリー時に入力した声を再生することで、視覚に頼らずに選出者を把握することができる。そのため、端末の画面が小さかったり参加者が大人数であったりしても結果をすぐに共有でき、作業中の人や視覚障がい者も参加しやすい抽選システムとなっている。

4.2 声によるエントリー方法

Facelot は顔をエントリーとして使用している。これは顔が名前を知らずともあるグループ内では個人を特定するた

めのキーになることで、事前に抽選のための準備を省力化している。また顔は誰であるかを判断する一般的な方法であるため、顔をキーとすることは見ただけでもすぐに誰が抽選されたかを認識しやすい。一方で Voicelot は人それぞれ声異なることをエントリーとして使うことで個人の特定を実現している。しかしながら、顔とは異なり声だけを聞いてもその声为谁であるかを参加者が把握することは親しい友人間でないとなれば難しい場合がある。また参加者数が増えればこの問題はより難化する。Facelot においても顔の写真がうまく取得できないことにはエントリーが困難であるため、顔と声のどちらを用いるかは、利用状況に応じた問題である。

5. 今後の展望

現在はエントリーを一人ずつ順に行っているため、エントリーが終了するまでに掛かる時間は参加者数に比例して長くなってしまふ。機械学習のアプローチにより、複数人が同時に発話した音を分離する技術の研究が進んでいる。今後この技術を用いて更なる時間短縮を検討している。

抽選の際に、声紋認証を用いることが可能である。声紋認証を用いる場合、声以外の個人情報結びつけることができるが、声の事前登録や個人情報の事前入力が必要になるため、トレードオフの関係にある。

参考文献

- 1) 伊藤暁, et al. ジャンケンの計算量. 電子情報通信学会論文誌 D 2003, Vol. 86, No. 7, pp. 452-457, 2003.
- 2) 金子翔麻, 渡邊恵太. Facelot: 顔検出と顔属性をエントリーとしたアドホックな抽選システム. エンターテイメントコンピューティングシンポジウム 2018 論文集, Vol. 2018, pp. 211-212. 2018.

‡ Web Empath API <https://webempath.net/lp-jpn/>

§ Google Cloud Speech-to-Text API <https://cloud.google.com/speech-to-text>