

音楽理解技術を利用したマイクロタスク埋め込み型 音楽ゲームの自動生成

中村 聡史^{†1 †2} 十文字 優斗^{†3}

世の中には人間による答えを必要とするタスクが膨大にあるが、そのタスクをこなすにはかなりの時間を要するものが多く、ある特定のユーザに処理を依頼するのは現実的ではない。こうしたタスクを数秒で回答可能なマイクロタスクに細分化し、多くの人に協力してもらうことで元のタスクを達成することを目的とした研究が多数なされている。このマイクロタスクへの回答を促すため、我々は音楽ゲームにマイクロタスクを埋め込む仕組みを提案してきた。しかし、これまでのシステムでは音楽ゲームをその曲ごとに作成する必要があったため適用可能な曲が限定的であった。本研究では、音楽理解技術によりマイクロタスクのための音楽ゲームを自動生成する手法を提案し、任意の動画に対して音楽ゲームを楽しみつつマイクロタスクに協力可能な仕組みを実現する。

1. はじめに

ある写真に写っている動物について、コンピュータがその動物が犬なのかそうでないかを判断することは容易ではなく、スキャンして取り込んだ文字の認識結果が正しいかどうかを判断することはコンピュータ自身にはできない。だが、人であればその問題に答えることは容易である。また、ある2つの手書き文字が提示されており、どちらが美しい手書き文字であるかということもコンピュータにはできない。こうしたコンピュータでは答えることが難しい、またはコンピュータが答えることが出来ないが、人であれば容易に答えることの出来る問題は世の中に多数ある。

コンピュータが処理することが難しいものについては、人にアノテーションを要求することになるが、そのアノテーション量が膨大な場合、少ない人数でこなすことは困難である。そうした際に、膨大なタスクを細かいタスクに細分化し、多くの人に分担して作業を要求する Human-based Computation [5] にまつわる研究は多数なされており、多くのシステムが実現されている。特に、ウェブ上の各種の認証で用いられている CAPTCHA を、画像上のテキストにどのような文字が書かれているのかを認識することに活用する reCAPTCHA[7]は世界中で日々利用されている。また、Crowd4U[3][4]は、膨大なデータをマイクロタスクに分割し、ウェブサイトでファイルをダウンロードする際にマイクロタスクを表示し、マイクロタスクへの回答を促す仕組みを実現している。さらにその発展で、通路の床にマイクロタスクを提示し、その通路を通る人が少し左右に歩く方向を変えるだけで質問への回答を可能とする研究がなされており、生活空間の中でのマイクロタスクへの回答が可能になっている[12]。

以上のように、これまでの研究でも、様々な状況や場所においてマイクロタスクを提示し、ユーザに何気なく回答

を促すような研究がなされている。しかし、現時点でもマイクロタスクを行うことが可能な機会は限られているため、さらに新しい場へのマイクロタスクの埋め込み、多くのユーザをマイクロボランティアに巻き込む仕組みを実現することが重要である。

ここで我々は、この十数年近く流行を見せており、いまだ勢いが衰えない音楽ゲームに注目する。音楽ゲームとは、プレイヤーが音楽に合わせてタイミングよくボタン操作を行うゲームであり、その市場は売上高150億円ほどである[9]。音楽ゲームではボタン操作のタイミングがどの程度合っていたのかということに応じてスコアが決まる。我々は、この音楽ゲーム中のタイミングにマイクロタスクを埋め込み、ボタン操作によって埋め込まれたタスクへ回答可能とする事で、マイクロタスクへの回答を集める手法を提案してきた[17]。しかし、これまでの手法では音楽ゲームの譜面を、我々が構築した譜面生成ツールを用いて事前に用意する必要があり、音楽のバリエーションを増やすことが困難であった。また、マイクロタスクも自身のサーバから取得するだけであるためタスクを追加する際に制約が多かった。

そこで本研究では、音響分析技術を用いて音楽ゲームの譜面を自動生成する手法を提案し、プロトタイプシステムを実装し、プロトタイプシステムを用いた検証によりその可能性や問題点について検討を行う。なお、プロトタイプシステムはウェブベースのシステムとして実装することにより、スマートフォン上でも動作するようにする。また、マイクロタスクの基盤システムである Crowd4U とも連携することにより、大規模なマイクロタスクへの貢献も行う。

2. 関連研究

マイクロタスクへの回答を集める手法としてこれまで様々な研究が行われてきた。

マイクロタスクの品質向上を目的とした研究として、タ

†1 明治大学
Meiji University
†2 JST CREST

†3 東京農工大学
Tokyo University of Agriculture and Technology

スクの割り当て手法の提案[15]や、タスクのインセンティブに関する研究[16][18]、タスクの質問文の改善に関する研究[14]など様々である。こうした研究を参考にすることにより、タスク結果の品質向上が可能であると考えられる。我々の研究でも、こうした過去の研究を参考に、質問文の改善などを行っていく予定である。

Amazon Mechanical Turk[2]やYahoo!クラウドソーシング[8]では、Webサイト上で提示された選択肢から当てはまるものをクリック、または文章を書くなど、複数の手法でタスクに回答を可能とするものである。本研究では、タスクへの回答方法は一つであり、選択肢を二つに限定することで、ユーザがタスクについて深く考えることなく、より手軽なタスクへの回答を可能とするものである。

マイクロタスクへの回答方法についての研究として、[12]では、床に掲示したタスクの上を通り抜けた時の情報を利用してタスクの回答を可能としている。本研究では、これとは異なり、よりユーザ本来の目的を阻害しない仕組みを用いてタスクへの回答を可能とするものである。本研究で提案する手法は、音楽ゲームという新しい場にマイクロタスクへの回答機会を設けるものである。

クラウドソーシングで集めたデータをもとにユーザを支援する研究は様々なものが行われている。例えば[19]では、コミュニケーションを支援するため、応答用例対を収集する方法を提案している。また、[1]では、写真に対する手描きなぞりのストロークを蓄積し、その蓄積したストロークから、ユーザの手描きを修正する手法を提案している。今後、このようなマイクロタスクやクラウドソーシング系のシステムは増えていくと考えられる。

3. 音楽ゲームと従来手法

音楽ゲームでは、どのボタンをどのタイミングで押すのかということが、そのタイミングの数秒前から画面上にマークとして提示されている。このマークのことを「ノート」と呼び、プレイヤーは画面上に表示されているノートに従いボタン(どのボタンを押すか示されている)を押していく。ここで、基本的にそのタイミングで押すボタンは1つに決まっている。

一般的に音楽ゲームでは、このノートが画面内のターゲット上に移動してきたタイミングでボタンを押すことが出来るとスコアが入る。また、連続的にタイミングに合わせてボタンを押すことができている場合、コンボ数が増えていき、そのコンボ数に応じてスコアの増加量が上昇していく。一方、タイミングと外れている場合は、コンボが0に戻り、タイミングから外れている数が多い場合に、ゲームオーバーとなる。

我々は、マイクロタスクをノートと同じように移動させ、ノートと同じようにボタン操作でマイクロタスクへの回答

を可能とした[17]。ここで、音楽ゲームに溶け込んでいるマイクロタスクが自然な形で選択問題として提示されると、プレイヤーはゲーム的には正しい回答を選ばなくても良いのに、間違った回答を選びたくないという気持ちから、正しい回答を選びたくなるのではないかと考えた。そこで、押すボタンが1つに割り当てられる単純ノートに加え、複数のボタンにより選択入力を実現する選択ノートを新たに導入した。この選択ノートでは、選択肢としてマイクロタスクを提示し、マイクロタスクに対する回答を選択ノートにおける各ボタンに割り当てる。この仕組みにより、プレイヤーは音楽ゲーム中のマイクロタスクへの回答が可能となっていた(図1)。

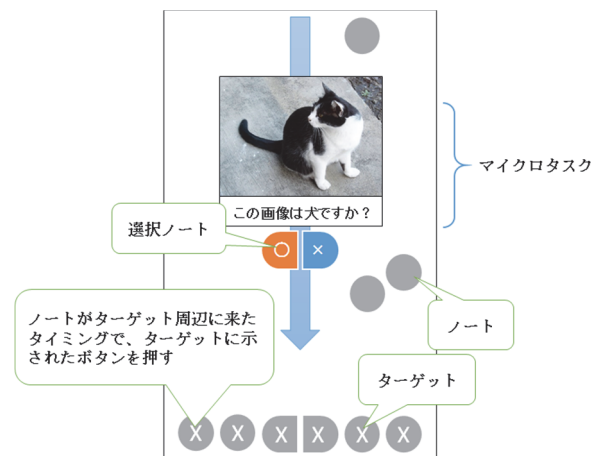


図1. 単純ノートと選択ノート

なおプレイヤーは、マイクロタスクに対して正しく回答しても、そのマイクロタスクの選択肢を無視して割り当てられているどちらかのボタンを適当に押しても、タイミングよくボタンをおすことができれば、コンボを増やしていくことを可能とした。つまり、プレイヤーは完全にマイクロタスクを無視することも可能となっている。しかし、従来手法では音楽ゲームの譜面を事前に用意する必要があり、すぐに飽きてしまうという問題があった。

4. 提案手法

音楽ゲームの譜面の自動生成において重要となるのは、生成する音楽ゲームの譜面の難易度が適切であるかということ、音楽を演奏しているかのような一体感や気持ちよさがあること、音楽ゲームをよりうまくプレイするための熟達の可能性があることなど様々である。しかし、そうした点をすべて盛り込むことは容易ではない。

そこで本研究では、まずすべての要素を盛り込むのではなく、音楽ゲームとして破綻しないことや、ある程度飽きずにプレイできることなどを考慮して、音楽ゲームのための譜面を生成する。なお、譜面を生成するというのは、楽曲動画の適切な生成時間に対してノートを挿入し、ボタンの割り当てを行うということである。

ここで音楽ゲームでは、音楽に対してリズムをとって楽しむことができるという事が重要である。そのため、ドラムなどのビートやメロディの音のタイミングが重要であると考えられる。また、音楽が無い部分は重要ではなく、盛り上がりが出されているサビ部分は重要であると考えられる。そこで音楽をある種の構造的なものと考え、「音が無いところ」「サビ部分」「サビ部分以外」に分類し、下記の点について考慮する。

- 音が鳴っていないところはノートを生成しない
- サビはメロディが重要である事が多いため、サビではメロディにあわせてノートを生成する
- サビでない部分については、ビートのパターンを重視し、ビートにあわせてノートを生成する
- 音楽の盛り上がる部分では、ノートを多数生成する

上記を実現するため、後藤らの音楽理解技術[10]にもとづくサビ区間検出手法が実現されている Songle[6][11]のAPI から、サビか否かという曲の構造情報と、ビートパターン、メロディーラインといった曲自体の詳細情報を取得する。また、下記のアルゴリズムに従い、まずは基本的な譜面を生成する。

1. 時間 $t=0$ とする
2. t 秒目に音がある場合は 3 へ、ない場合は 6 へ
3. t 秒目がサビの場合は 4 へ、サビでない場合は 5 へ
4. t 秒と $t+\Delta t$ 秒の間にメロディの音に変化している場合はノートを挿入、そうでない場合は 6 へ
5. t 秒がビートの開始位置の場合はノートを挿入、そうでない場合は 6 へ
6. t に Δt を加え、2 に戻る

次に、盛り上がりによりノートの数を増加させる場所を決定する。ここでは、ニコニコ動画において動画の再生時間軸に付与されたコメントの量からその盛り上がりを取得し、盛り上がっている部分のノート数を倍増させる。盛り上がっている場所はコメントの総量で正規化し、その値がある一定以上を超えている部分とする。なお、盛り上がっている部分へのノートの増加については、ここでは単純にノートとノートの間に挿入していくことで倍増させている。

最後に、選択ノートを挿入する位置を決定する。また、 t_s 秒に 1 度のペースで、単純ノートを選び、選択ノートへと差し替える。ここで、マイクロタスクの手前に一般ノートが集中すると、マイクロタスクへの回答を考える余裕がなくなる。また、選択ノートとセットで移動してくるマイクロタスク自体に一般ノートがかぶるとマイクロタスクを判別できなくなるという問題がある。そこで、選択ノートの前後のノートを削除することにより、音楽ゲームプレイの邪魔にならないようにする。

5. プロトタイプシステム

プロトタイプシステムは、ウェブシステムとして JavaScript と PHP を用いて実装した。また、jQuery をライブラリとして利用し、音響分析結果の取得に Songle API を、マイクロタスクの取得に Crowd4U の API を利用した。システムの動作例は図の通りである。

図 2 は、システムの起動画面である。ユーザは Songle で音響認識されているすべての楽曲を音楽ゲームとしてプレイすることが可能となっている。また、図 3 は実際にプレイしている様子である。PC 版では画面左側に Songle 自体のプレイヤーが、右側に音楽ゲームが提示され、Songle プレイヤの楽曲動画再生にあわせて音楽ゲームのノートが上から下に移動してくる。ここでゲームのプレイヤーは、ノートが画面下のターゲットの位置に来た時にタイミング良くボタンをおすことで、高得点を獲得することが可能となっている。



図 2. 起動画面。人気の曲が表示されており、テキストボックスから曲を検索することが可能



図 3. PC 版でプレイしている様子

図4および図5はシステム上でマイクロタスクが提示されている様子である。図4は平均手書き文字が美しいかどうかを人から判断されるのかどうか[13]ということを検証するためのマイクロタスクが提示されている。また、図5ではどの程度ユーザがマイクロタスクに真剣に答えようとしているのか、どの程度ユーザはマイクロタスクに回答することができるのかを確認するために差し込まれている、誰もが同じ回答が可能であるマクロタスクの提示例である。

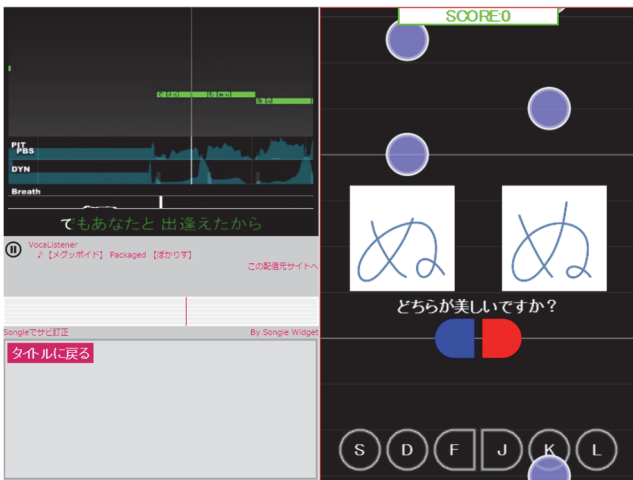


図4. マイクロタスク（どちらが美しいと感じるかを問うタスク）が提示されている様子



図5. マイクロタスク（大きい数字を選択するタスク）が提示されている様子

プロトタイプシステムを iPhone でプレイしている様子は図6および図7の通りである。iPhone 上では、Songle のプレイヤーは画面上部に提示される。また、キーを押すのではなく、ノートが画面下部のターゲットが部分に移動してきたタイミングにあわせてターゲットの部分で両手の指でタップするようになっている。



図6. iPhone 上での動作例

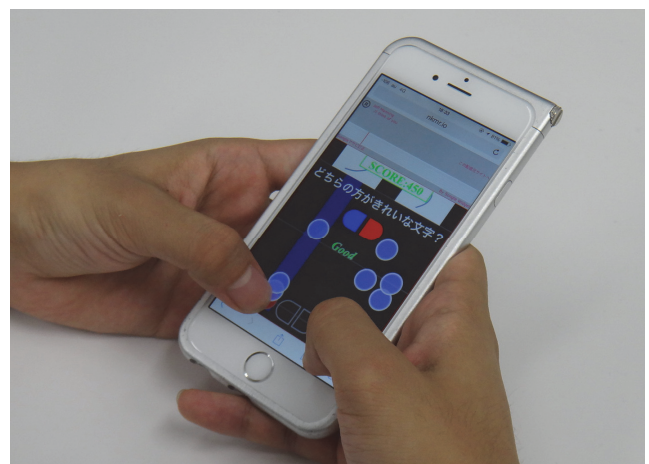


図7. iPhone 上でプレイしている様子

6. フィードバックと考察

音響理解技術を用いた音楽ゲームの譜面生成の精度について、実際に複数名にプレイしてもらうことにより、その良さや問題点などを検証した。

まず、Songle API による音楽の理解技術によって、膨大な楽曲をカバーできているため、自分の好みの音楽をゲームとして楽しむことができ、長く楽しむことができるという感想を得た。実際にメジャーなアーティスト名や楽曲名、初音ミクなどの VOCALOID で作成された有名曲などが検索され、音楽ゲームとしてプレイされていた。自身が好きな楽曲を音楽ゲームとしてプレイできるというのはかなり楽しいものであったらしく、評判が良かった。この任意の楽曲動画を音楽ゲームとしてプレイできるというのは狙っていたことであり、この点については目的が達成できたといえる。また、音楽ゲームの譜面に気持ち悪さを感じているというフィードバックが無かったことから、生成された譜面が音楽ゲームとしてそこまで大きな問題でなかったことがわかる。一方、バラードなどのゆっくりしたテンポの

楽曲では、ゲームの難易度が高いという報告があった。これは、音楽ゲーム特有の問題であると考えられるため、なかなか対応が難しいものである。

生成された譜面に関するフィードバックとしては、ゲームの難易度に関するものが多かった。今回の手法では、単位時間あたりのビートの量や、メロディの音の量が多い曲においては単位時間あたりのノート数が増加し、量が少ない場合はノート数が減少することになる。そのため、楽曲によって難易度が上がったり下がったりしていた。ここで音楽ゲームに慣れているユーザは、難易度が高い譜面を楽しくプレイすることができる（現システムの高難易度のものよりさらに難易度が高いものや、速度を速めるハイスピードモードが望まれていた）一方で、難易度が低い譜面をあまり楽しくないと感じ、音楽ゲームに不慣れたユーザは難易度が高い譜面だとタイミング良くボタンを押すことが出来ず、ただただ焦ってしまうということになっていた。こうした問題については、譜面生成においてレベル設定を可能とする手法を実現する必要がある。

今回の音楽ゲームでは「D」「F」「J」「K」という4つのキーで単純ノートを、「G」「H」という2つのキーで選択ノートをタイミング良く押すよう割り当てていた。ここで、譜面として「D」や「K」を押すものが多く、「F」や「J」を押すものが少なかったため、もっと単純ノートのキー配置が偏りなく出てきて欲しいというコメントや、同時押しなどの操作が欲しいなどのコメントがあった。これについては、どういった配置が音楽ゲームとして楽しいのか、どういった時に同時押しが良いのかなどを考えつつ改良していく予定である。また、今回は音楽のビートやメロディを単純にキーに割り当てているだけであり、どういったビートの連続性、メロディの連続性の時にどういったノートを生成すべきなのかといったことが一切考慮できていなかった。こうした問題については今後の研究で改善していく予定である。

一方、今回の手法では認識されたビートパターンを利用していたため、音楽がない喋り声の部分がビートとして抽出されてしまっている楽曲では、その部分にノートが設定されてしまっていた。そのため、ただの喋り声の部分を音楽ゲームとしてプレイする必要があり、苦痛であるなどの問題があった。この問題については、ビートとして検出されている部分が音楽か否かという判定を行うことにより改善可能であると考えられる。

次に、マイクロタスクをどの程度回答可能であったかという点については、ほとんどのユーザが、最初の数回ではマイクロタスクにまともに回答することが出来ていなかった。プレイしている様子を観察していると、初期段階では多くのユーザが、マイクロタスクが出た時に慌てて両方のボタンを押すなどしていた。一方、マイクロタスクにある程度慣れ、どのような問題が出てくるのかというパターン

を理解していくと、「こういった画像が出てきた時はこういうタスクだ」と判断して、瞬間的に判断できるようになっていた。このことから、慣れによりマイクロタスクは十分にこなせるものと考えられる。しかし、一部のユーザから、慣れてきても「どちらの数字が大きいか？」を問うタスク（図5）については、即座に回答できないというフィードバックがあった。これは、問題自体を英語で提示していたことと、タスク自体が小さく提示されており、その数値の判別が難しいことが理由であることが考えられる。今後はこうした問題を改善していき、本質的な問題は何かということをも明らかにする予定である。

なお、マイクロタスクを間違ってしまう理由については、同じ音楽ゲーム内で様々なマイクロタスクが提示されてしまい、状況に応じて問題を読み、判断する必要があると考えられる。同一のマイクロタスクであれば、特に深く考える必要はなく、瞬間的に選択可能になると考えられる。この検証については今後の課題である。

その他のフィードバックとしては、タイミングがずれて気持ち悪いと感じる部分があるというものであった。これは、ウェブブラウザ上で動画を再生しながらJavaScriptを使って音楽ゲームを実現しているため、環境によっては処理がもたつくことが原因であった。音楽ゲームにおいてタイミングは非常にシビアなものであり、微妙な遅れも気になってしまうものである。こうした点については、処理速度を考慮して表示の軽量化などを行うことで対応する予定である。

最後に、長時間のプレイは飽きてしまうという感想もあった。現在のシステムは楽曲全体を音楽ゲームとしているため、どうしてもこのような問題は生じてしまう。多くの音楽ゲームでは、長くて2分程度に設定されている。この問題については、サビの認識を行い、途中を飛ばすなど工夫することによって改善する予定である。

7. おわりに

本研究では、マイクロタスクを埋め込む場面として音楽ゲームを検討し、その音楽ゲームの譜面を、音楽理解技術を応用して自動生成する手法を実現した。この手法により、任意の楽曲に対して音楽ゲームが自動生成でき、好みの楽曲を音楽ゲームとして楽しむことができるようになった。また、この音楽ゲームに対して埋め込まれたマイクロタスクについても、ある程度慣れることによりこなすことができるようになることがわかった。

今後の課題は、ゲームの難易度を設定可能とすること。また、マイクロタスクを問題なく回答可能とするために、同じ楽曲内ではマイクロタスクの問題を統一することなどが考えられる。さらに、音楽ゲームとしての面白さをより引き出すための仕組みを検討する。

謝辞 音楽分析技術のための Songle API を整備および提供いただいている産業技術総合研究所の後藤真孝様および、Crowd4U の仕組みを実現し API の提供を頂いている筑波大学の森嶋厚行先生に感謝致します。

参考文献

- 1) Alex Limpaecher, Nicolas Feltman, Adrien Treuille, Michael Cohen: Real-time drawing assistance through crowdsourcing, ACM SIGGRAPH 2013 Conference Proceedings, Volume 32 Issue 4, No. 54 (2013).
- 2) Amazon Mechanical Turk, <https://www.mturk.com>
- 3) Atsuyuki Morishima: Cylog/crowd4u: A case study of a computing platform for cybernetic dataspace, In Pietro Michelucci, editor, Handbook of Human Computation, Springer, 2013.
- 4) Crowd4U, <http://crowd4u.org/>
- 5) Edith Law and Luis von Ahn: Human Computation. Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Publishers, 2011.
- 6) Songle, <http://songle.jp/> (2014).
- 7) von Ahn, L., Maurer, B., McMillen, C., Abraham, D. and Blum, M. reCAPTCHA: Human-Based Character Recognition via Web Security Measures, Science 321, 5895, pp.1465-1468 (2008).
- 8) Yahoo! クラウドソーシング, <http://crowdsourcing.yahoo.co.jp/>
- 9) 一般社団法人 全日本アミューズメント施設営業者協会連合会, 平成 24 年度アミューズメント産業界の実態調査, <http://www.aou.or.jp/news/1310/06.pdf>
- 10) 後藤真孝, SmartMusicKIOSK: サビ出し機能付き音楽試聴機, 情報処理学会論文誌, Vol. 44, No. 11, pp. 2737-2747 (2003).
- 11) 後藤真孝, 吉井和佳, 藤原弘将, Matthias Mauch, 中野倫靖: Songle: 音楽音響信号理解技術とユーザによる誤り訂正に基づく能動的音楽鑑賞サービス, 情報処理学会論文誌, Vol.54, No.4, pp.1363-1372 (2013).
- 12) 品川有輝, 森嶋厚行, 中村聡史, 寺田努: 日常空間に組み込んだ Human Computation 環境によるクラウドソーシングタスク処理, 情報処理学会インタラクション 2014, pp.706-707 (2014).
- 13) 中村聡史, 鈴木正明, 小松孝徳: 平均文字は美しい, エンタテインメントコンピューティングシンポジウム 2014 論文集, 32-39 (2014).
- 14) 丹治寛佳, 清水伸幸, 森嶋厚行, 北川博之: マイクロタスク型クラウドソーシングにおける質問文改善の支援手法, DEIM Forum 2015, C6-1 (2014).
- 15) 巻口誉宗, 東正造, 下村道夫, 金丸直義: ヒューマンコンピューテーションにおけるタスク割り当て手法の提案, JSAI2014 オーガナイズドセッション, pp. 1-4 (2014).
- 16) 松原繁夫, 伊奈祐輔: クラウドソーシングでのタスク選択に関する行動モデリング, JSAI2014 オーガナイズドセッション, 1J4-OS-18a-3, pp. 1-4 (2014).
- 17) 三輪聡哉, 中村聡史: マイクロタスク埋め込み型音楽ゲームの提案, 第 34 回情報処理学会エンタテインメントコンピューティング研究会, 2014-EC-34(2), pp.1-6 (2014).
- 18) 堀田裕理, 松原繁夫: 成果報酬によるクラウドワーカーの制御, JSAI2014 オーガナイズドセッション, 1J4-OS-18a-5, pp.1-4 (2014).
- 19) 山本里美, 福島拓, 吉野孝: クラウドソーシングにおける会話をういた応答用例対作成手法の提案, 情報処理学会論文誌, Vol.56, No.3, pp.1080-1089 (2015).