

SNS ストリームの効率的な「聴き飛ばし」を実現する 音声ブラウジング手法

岩永 章吾^{1,a)} 牛尼 剛聡^{2,b)} 角谷 和俊^{3,c)}

受付日 2015年3月21日, 採録日 2015年7月10日

概要: 近年, 爆発的に普及した SNS では, コンテンツが一系列に並んだストリーム (SNS ストリーム) としてユーザに配信される. 一般的に, SNS ユーザは, 配信される SNS ストリームを構成するすべてのコンテンツを読むわけではない. ユーザは, 個々のコンテンツの一部を見て, そのコンテンツのユーザにとっての価値を推定し, 価値が低いと推定されたコンテンツは「読み飛ばし」を行う. 視覚が利用できない状況や, 視覚障害者がコンテンツを閲覧するために, 音声ブラウジングが広く利用されているが, SNS ストリームの音声ブラウジングでは, それを構成するすべてのコンテンツを読み上げるため, 効率的な「聴き飛ばし」が困難であるという問題点がある. 本論文では, SNS ストリームの効率的な音声ブラウジングを実現するために, ユーザの効果的な「聴き飛ばし」を支援する手法を提案する. 本手法では, 個々のコンテンツの投稿者とユーザとの関係性が, そのコンテンツの価値を推定するための重要な手がかりになるという仮説の下に, 投稿者をユーザのフォロー関係に基づいてグループ化し, グループごとに異なる位置 (定位) から読み上げられていると感じられるように読み上げ音声を再生する. Twitter を対象に提案手法を実装したプロトタイプシステム Audio Twipp を利用した評価実験の結果から, 提案手法は, すべてのツイートを同一の位置から読み上げた場合と比較して, ユーザが「聴き飛ばし」を行うまでの時間が短縮化され, 一定時間内により多くの読み飛ばしが行われるようになり, 結果的に同一時間内により多くの Twitter を聴取可能であることが明らかになった.

キーワード: SNS, 音声ブラウジング, 読み飛ばし, SNS ストリーム

An Audio Browsing Method for Effective Skip Listing of an SNS Stream

SHOGO IWANAGA^{1,a)} TAKETOSHI USHIAMA^{2,b)} KAZUTOSHI SUMIYA^{3,c)}

Received: March 21, 2015, Accepted: July 10, 2015

Abstract: Today, in widely spread Social Networking Services (SNSs), contents are delivered as a stream (SNS stream) to users. Generally, SNS users would not read each content in an SNS stream perfectly. Users would estimate the value of each content for themselves based on some parts of each content and they skip the content if it is judged less valuable. Audio browsing is widely used in situations where users cannot their vision and where visually impaired persons brows contents. However, when users brows an SNS stream with conventional audio browser, every content in the SNS stream have to be read aloud perfectly. Therefore, it is difficult for the users to brows the SNS stream effectively. For effective audio browsing of an SNS stream, this paper proposes a method for supporting effective “skip listing.” In this method, users who have posted content in an SNS stream are grouped based on the assumption that relationships between users would be important cues for estimating the value of the content, and the audio of the content is played at the orientation of the group of the user. Based on the experimental studies with Audio Twipp, which is a prototype system, the proposed method enables users to brows contents more effectively in an SNS stream than conventional simple audio browsing.

Keywords: SNS, audio browsing, skip reading, SNS streams

¹ 九州大学大学院芸術工学府
Graduate School of Design, Kyushu University, Fukuoka
815-8540, Japan

² 九州大学大学院芸術工学研究院
Faculty of Design, Kyushu University, Fukuoka 815-8540,
Japan

³ 関西学院大学総合政策学部
School of Policy Studies, Kwansai Gakuin University, Sanda,

1. はじめに

近年, Facebook や Twitter に代表されるソーシャル・

Hyogo 669-1337, Japan

a) s.iwanaga.171@gmail.com

b) ushiama@design.kyushu-u.ac.jp

c) sumiya@kwansai.ac.jp

ネットワーキング・サービス (SNS) が爆発的に普及した。SNS 上には、継続的に新しいコンテンツが投稿され、ユーザには、あらかじめ指定した他のユーザが投稿したコンテンツが配信される。現在、SNS は情報獲得やコミュニケーション等、様々な目的のために使用されている。

一方で従来から情報獲得のためには Web 検索が広く利用されてきた。Web 検索では、ユーザは検索要求をクエリとして検索エンジンに与え、検索エンジンは与えられたクエリに合致する Web ページを検索結果として返す。Web 上に存在する膨大な情報から、ユーザが必要な情報を見つけるために、検索エンジンは重要な役割を果たしてきた。

一般に、Web 検索による情報獲得は、ユーザが獲得したい情報を具体的にイメージできることが必要であり、特定の製品の仕様を知りたいといった、短期的な情報要求には適している。しかし、継続的に特定の分野に関して価値がある情報を取得したいという発見的な検索要求には対応できない。SNS では、ユーザは自分にとって価値が高い情報を提供してくれる可能性が高いユーザが投稿する情報を継続的に受信する。これにより、SNS は、受信した情報の中からユーザに対して、新規性がある発見的な情報を取得したいという要求に適していると考えられる。

SNS では、ユーザにとって価値が高い情報が配信されることがあるが、配信される情報すべての価値が高いわけではない。ユーザは、自分に配信された情報の中から、自分の情報要求に合致する情報を手動でフィルタリングする必要がある。SNS では、ユーザは、自分に配信されたすべてのコンテンツの内容を正確に把握するわけではない。ユーザは、配信された多数のコンテンツに対して、自分にとって価値が高いと推定されるコンテンツに対しては、正確にその内容を把握する一方で、自分にとって価値が低いと推定されるコンテンツに対しては、その内容を正確に把握しないことが多い [1]。このように、配信されたコンテンツに対して、そのコンテンツの属性や本文の一部のみから、ユーザにとっての価値を推定し、価値が低いと考えられるコンテンツに対しては正確な内容把握を行わない行為を「読み飛ばし」と呼ぶ。

SNS における情報取得において効率的な「読み飛ばし」を実現することは重要である。そのために、SNS では、ユーザに配信されたコンテンツは一列に直列化され、ストリーム形式でユーザに提示されることが多い。本論文では、このようにストリーム形式に構成された SNS コンテンツを SNS ストリームと呼ぶ。Twitter のタイムラインは SNS ストリームの代表例である。SNS ストリームは、パーソナルコンピュータ上でのマウスのホイールやスマートフォン上でのフリック操作を利用したスクロールにより、単純な操作で効率的に「読み飛ばし」可能である。

一方、スマートフォンやタブレット型コンピュータ等の携帯型情報端末の普及により、ユーザは「いつでも」「ど

こでも」SNS ストリームをブラウジングできるようになった。それとともに、電車やバスの待ち時間等、日常生活のちょっとした隙間時間を利用して、SNS ストリームをブラウジングすることが多くなった。現在、SNS ストリームのブラウジングのためには、携帯型情報端末の画面上に表示されたコンテンツを視覚を利用して閲覧する手法が一般的である。しかし、歩行中や自動車の運転中に、視覚を利用する SNS ストリームの閲覧行為を行うことは危険である。また、料理や掃除等の家事を行いながら、視覚を利用しない手法で SNS ストリームを閲覧したい場合もある。また、視覚障害者や高齢者等はスマートフォンの小さな画面では、視覚を利用した閲覧が困難な場合がある。こうした中で、視覚を利用しないブラウジング手法として、音声ブラウジングの重要性が高まっている。

SNS ストリームの音声ブラウジングを行う場合、従来の一般的な音声ブラウジング手法では、SNS ストリームを構成する 1 つ 1 つのコンテンツに含まれるテキスト要素をすべて読み上げる。この手法では、視覚ブラウジングよりも、1 つ 1 つのコンテンツの価値を推定するために時間がかかり、効率的な「読み飛ばし」を実現することが困難である。我々は、音声ブラウジングにおける「読み飛ばし」を「聴き飛ばし」と呼ぶ。本論文では、視覚ブラウジングにおける「読み飛ばし」の特徴を分析し、その結果に基づいて効率的な「聴き飛ばし」を実現するための手法を提案する。

Twitter をはじめとする多くの SNS では、視覚ブラウジングを行う際に、SNS ストリームを構成する個々のコンテンツに投稿者を区別するための画像アイコンを表示している。一般に、画像情報は文字情報よりも直感的かつ効率的に区別が可能であるため、コンテンツに投稿者の画像アイコンが含まれていることで、ユーザはコンテンツの投稿者を効率的に判断できると考えられる。SNS ストリームに含まれる個々のコンテンツを「読み飛ばす」か否かを決定するために、ユーザは、コンテンツに含まれる一部分の情報から、ユーザにとって、そのコンテンツがどの程度の価値があるかを推定する必要がある。我々は、この推定を行う際に、対象とするコンテンツの投稿者を把握することが重要な役割を果たしていると考えた。なぜなら、投稿者によって、投稿されるコンテンツの種類や内容がある程度限定できるため、投稿者を把握できれば、本文の一部を見ただけでも、そのコンテンツの価値を効率的に推定可能となると考えられるからである。従来の SNS の音声ブラウジング手法においては、画像アイコンのように、投稿者をユーザに効率的に伝達する手段を提供していない。したがって、音声ブラウジングにおいても、ユーザにコンテンツの投稿者を効率的に伝達することができれば、効率的な「聴き飛ばし」を実現できると考えられる。

本研究では、音声ブラウジングにおいて、投稿者を区別するために、読み上げ音の定位の違いを利用するアプロー

チを利用する。定位とは、音を聴取した人物が、その音が発せられていると感じる位置のことである。我々が日常的に利用しているステレオヘッドホンを利用した音再生環境では、再生音の左右の音量バランスを変化させることで、比較的簡単に定位を変更することができる。そこで、コンテンツの投稿者によって、決められた定位で読み上げ音を再生することで、ユーザが効率的に投稿者を区別できれば、効率的な「聴き飛ばし」につながると考えられる。しかし、一般に1人のユーザのSNSストリームを構成するコンテンツの投稿者は数十～数百であり、すべての投稿者を定位で区別することは困難である。

この問題を解決するために、我々は、投稿者をグループ化して、グループによって定められた定位で読み上げ音を提示することで、音声ブラウジングにおける効率的な「聴き飛ばし」を支援する手法を提案する。これは、個々の投稿者を区別することができなくても、ユーザと投稿者のおおよその関係性が把握できれば、コンテンツの価値推定の効率化につながるという考えに基づいている。本手法では、Twitterを対象として、フォロー関係から投稿者を自動的にグループ化し、グループごとに音の定位を変化させることで、ユーザが短時間で「聴き飛ばし」をするか否かを決定可能とする。さらに、本手法に基づいて開発したTwitter読み上げクライアント「Audio Twipp」を利用した被験者実験により、提案手法の有効性を評価する。

本論文は以下のように構成される。2章では関連研究を示す。3章では、SNSストリームの「読み飛ばし」の特徴を分析する。4章では、提案手法の具体的な内容について述べる。5章では、プロトタイプシステムについて述べる。6章では、評価実験の結果を示し、提案手法の有用性を評価する。7章では、まとめと今後の課題について述べる。

2. 関連研究

これまでも、音声ブラウジングに関する様々な研究が行われている。

海老名ら [2] は、視覚障害者のコンピュータ利用支援の手法において、マウスカーソルの位置とメニュー音声の位置の関係をを用いた手法を提案している。この手法では、たとえば、メニュー音声のマウスよりも左側にある場合、ユーザは左側に音像を感じるように音声フィードバックを提示することで、聴覚障害者が効率的なポインティング操作を支援する。Brewsterら [3] は、3次元音響空間を用いて、ユーザに対してメニューの提示と選択を行う手法を提案している。この手法では、ユーザが選択可能なメニューアイテムが3次元音響空間の異なる定位から音声イヤホンとして再生される環境下で、ユーザが頭部を動かすジュースチャによってアイテムを選択可能である。Sawhneyら [4] は、3次元音響空間の時計メタファを提案し、メッセージやリマインダメッセージを音声提示する際に、それらの時刻情報

を読み上げ音声の音像の定位で提示する手法を提案している。Gooseら [5] は、Webページのハイパーリンク構造を立体音響空間上で提示する手法を提案している。これらの手法では、音声ブラウジングにおけるユーザの効率的な操作を支援するために、音声ブラウジングにおいてユーザに提示する音像の定位の違いを利用している。

音声ブラウジングにおいて、効率的な情報取得を実現するために、カクテルパーティ効果を利用する手法が提案されている。河本ら [6] は、視覚障害者の効率的なコンピュータ利用を支援するために、GUI (Graphical User Interface) の視覚情報を音声に置き換える手法を提案している。この手法では、GUIの基本操作の1つであるメニュー選択に注目し、複数のメニュー項目を同時に読み上げることで、ユーザはカクテルパーティ効果により、メニューの項目を効率的に把握可能である。Schmandtら [7] は、カクテルパーティ効果を用いた音声ブラウジングツール AudioStreamer を提案している。このツールでは、物理空間において3カ所の固定された位置にスピーカを配置し、それぞれのスピーカから同時にニュースの音声データが再生される。再生中に、ユーザが特定のスピーカから再生されるニュースに興味を持った場合、それを選択すると、選択した音声データの音量が大きくなり、正確に内容を聴取可能となる。小林ら [8] は、単一の文書の効率的な音声ブラウジングを実現するための Dynamic Soundspace を提案している。このシステムでは、3次元音響空間上でのユーザを中心とする円形軌道上で、読み上げ音像の定位を一定速度で移動させながら再生する。複数の音源を同時に再生させることにより、カクテルパーティ効果により効率的な音声ブラウジングを実現することを目標としている。岩永ら [10] は、SNSストリームを対象に、それを構成する投稿を、3次元音響空間上の異なる定位から再生する際に、定位ごとに再生タイミングをずらすことで、読み上げ内容の聴きやすさと理解しやすさが向上することを示した。SNSストリームの音声ブラウジングを行う際に、複数の音声を同時に再生することは、一定の効果が期待できる。しかし、複数の音声を同時に聴取させることは、ユーザの負荷が増加し、煩雑に感じる場合があるという問題点がある。本論文では、複数の音像の同時再生によるカクテルパーティ効果を利用せず、効率的なSNSストリームのブラウジングを実現する手法を提案する。

3. SNSストリームを対象とした「読み飛ばし」の分析

3.1 SNSストリーム閲覧の特徴

ユーザはSNSストリームの閲覧時に、そのストリームに含まれるSNSコンテンツの内容をすべて正確に理解するわけではない。ユーザは、何らかの基準に基づいて、一部のSNSコンテンツのみ、内容を正確に把握し、残りに対

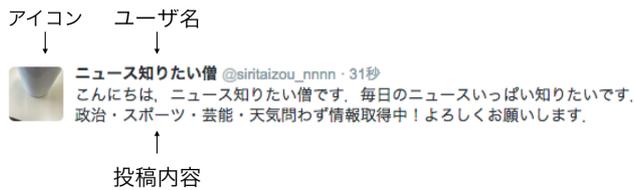


図 1 ツイートの構成

Fig. 1 The composition of a tweet.

しては、「読み飛ばし」を行うことが多い。本研究で対象とする音声ブラウジングにおいても、同様にすべての SNS コンテンツの内容を読み上げるのではなく、特定のコンテンツは正確な内容理解を行わない。本研究の目的は、ユーザが効率的に SNS コンテンツの「聴き飛ばし」を実現可能なブラウジング手法を提案することである。

音声ブラウジングによる効率的な「読み飛ばし」手法を開発するため、我々は、現在一般的に行われている視覚ブラウジングにおける「読み飛ばし」を分析した。視覚ブラウジングにおいて、1つのツイートは、図 1 に示すように、ユーザアイコン、ユーザ名、本文で構成される。SNS 閲覧の際、ユーザは、ストリーム内の個々のツイートに対して、ユーザアイコン、ユーザ名、本文に含まれる単語等から、その価値を推定し、「読み飛ばし」をするかを決定していると推測される。

ここで、SNS 上でのユーザアイコンは、各ユーザが固有のものを利用する 경우가多く、視覚的に直感的かつ効率的にユーザを区別可能である。そのため、読み飛ばしを行うとき、ユーザはアイコンと投稿内容から投稿者を判別し、「読み飛ばし」をするかを決定していると考えられる。そこで我々は、『タイムライン内のツイートを「読み飛ばし」をするかを判断する際に、ユーザアイコンが重要な手がかりとなっている』という仮説を立てた。この仮説を検証するために、次の実験を行った。

3.2 実験手法

ツイートに表示されるアイコンがある場合とアイコンがない場合で、ユーザが処理できるツイート数を調査する実験を行った。被験者は日常的に Twitter を利用している 9 人 (男性 7 人、女性 2 人) である。実験は、被験者に各自が日常的に利用しているスマートフォンを利用して、アイコンが見える場合とアイコンが見えない場合のそれぞれの場合で、30 秒間に閲覧したツイート数を調査することによって行った。具体的には、アイコンがない場合の閲覧を行う際には、スマートフォンの画面上でアイコンが表示される領域に紙を貼り、被験者からアイコンが見えないようにしてツイートを閲覧してもらった。図 2 はアイコンがある場合とアイコンがない場合のツイートの表示例である。ツイートの閲覧は、被験者が日常的に行っているのと同様なやり方で行ってもらうように依頼した。ツイート数は、



図 2 「アイコンがあるツイート」と「アイコンがないツイート」の例
Fig. 2 Examples of tweets with/without their icon images.

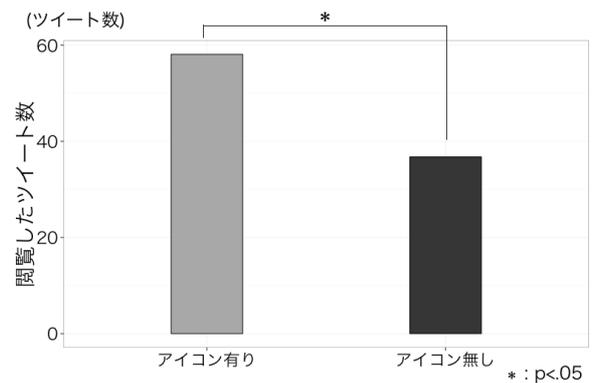


図 3 アイコンが表示される場合とされない場合に閲覧したツイート数の平均値

Fig. 3 The average numbers of tweets with/without icon images.

被験者のタイムライン上で実験開始前に表示されていたツイートと閲覧開始後 30 秒後に表示されていたツイートを記録し、その間に表示されたツイート数を目視により数えた。また、ツイートを閲覧するためのソフトウェアとしては、Twitter 公式クライアントを使用した。

3.3 実験結果

アイコンがある場合とない場合のツイート数の閲覧したツイート数の結果を図 3 に示す。アイコンがある場合、閲覧したツイート数の平均は、58.1 であった。一方、アイコンがない場合は、閲覧したツイート数の平均は、36.7 であった。また、それぞれのデータの分布が正規分布に従うという仮説をコルモゴロフ-スミルノフ検定を用いて検定を行った結果、アイコンがある場合は $p = 0.17 > 0.05$ であり、アイコンがない場合は $p = 0.31 > 0.05$ となり、ともに有意水準 5% で正規分布に従うという仮説を採択できる。そこで、対応のある t 検定で片側検定を行い、「アイコンがある場合はアイコンがない場合よりも多くのツイートを閲覧できる」という仮説を検証した結果、有意水準 5% ($p = 0.02$) でこの仮説が正しいことが示された。

さらに、同一の被験者に対するアイコンがある場合とアイコンがない場合の閲覧数の関係を調査するために回帰分析を行った。図 4 にデータの分布と分析によって得られた回帰直線を示す。横軸 (x 軸) はアイコンがない場合に閲

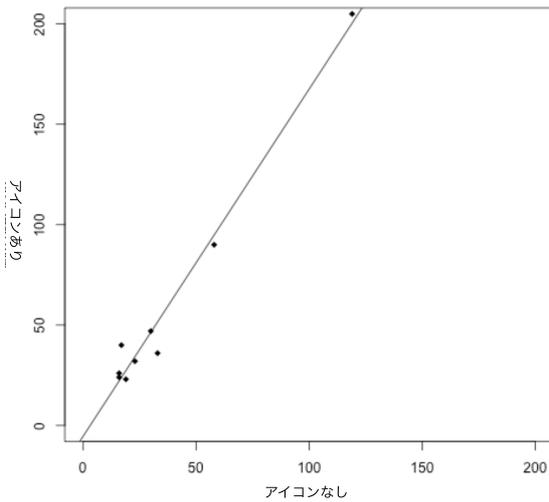


図 4 アイコンが表示される場合とされない場合に閲覧したユーザーごとのツイート数

Fig. 4 The relations that the number of tweets with/without their user icon images.

覧したツイート数であり、縦軸 (y 軸) はアイコンがある場合に閲覧したツイート数を表しており、点がそれぞれの被検者を表す。実線の直線が回帰直線を表し、得られた回帰直線の方程式は $y = 1.73x - 5.54$ であった。また、この回帰直線の寄与率は 0.98 であったため、この式は対応関係を非常によく表していると考えられる。このことから、アイコンがある場合は、アイコンがない場合よりも約 1.7 倍のツイートを閲覧可能であるといえる。

以上の結果から、アイコンがあることで、より多くの数のツイートを読むことができることが分かる。つまり、アイコンが「読み飛ばし」のための重要な手がかりになっていると考えられる。

4. 提案手法

4.1 アプローチ

3 章で示した実験により、視覚ブラウジングにおいては、ツイートの投稿者をアイコンを用いて直感的に分かりやすく表現することが、「読み飛ばし」を判断するために大きな影響を与えていることが分かった。SNS ストリームを音声ブラウジングする場合、ストリームを構成するコンテンツは一列に順番に読み上げられるため、ユーザーは投稿者を直感的に分かりやすく把握できない。そこで、音声ブラウジングにおいて、投稿者に関する情報を、視覚ブラウジングにおけるアイコンのように直感的に分かりやすく提示することができれば、音声ブラウジングにおいても閲覧の効率化が期待できる。

音声ブラウジングにおいて、投稿者情報を直感的に分かりやすく提示するために、提案手法では、読み上げ音の定位と投稿者情報を関連づけて提示する。ここで、音の定位とは、音源がどの位置にあるように感じられるかということ

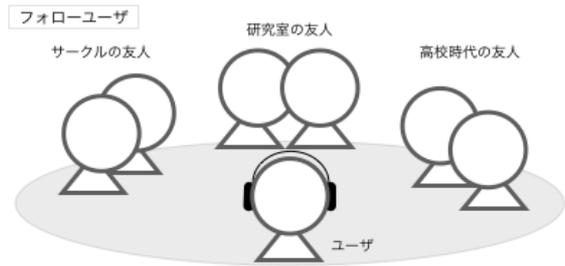


図 5 提案手法の概念図

Fig. 5 An overview of the proposal method.

である [9]。我々は、音の定位は、視覚ブラウジングにおけるアイコンと同様に、投稿者を直感的に区別可能ではないかと考えた。つまり、読み上げ音がどこから聴こえてくるかによって、投稿者が区別できれば、本文の読み上げが開始されてから早い段階で、そのツイートを最後まで聴取するか聴き飛ばすかを判断可能である可能性があると考えた。

しかし、一般に、1 人のユーザーがフォローするユーザー数は数十～数百であり、個々のユーザーを音の定位で区別することは現実的ではない。そこで、我々は投稿者をユーザーからの関連性に基づいてグループ化し、投稿者が所属するグループに応じて、決められた定位から読み上げ音声を再生する手法を提案する。投稿者のグループによって、話題の内容や種類はある程度限定できるため、本文の最初の数語を聴いただけで、読み飛ばすかどうかを判断できると期待できる。

提案手法は以下の 2 つのステップで構成される。

- (1) あらかじめ自分がフォローしているユーザーを Twitter のフォロー関係を利用してグループ化する。
- (2) SNS ストリームを順番に読み上げる。ここで、読み上げ音は、投稿者が所属するグループごとに定位を変化させて再生する。

図 5 に提案手法で読み上げを行う例を示す。図 5 では、左からサークルの友人の投稿、右からは高校の友人の投稿、中央からは研究室のフォロワの投稿が再生される様子を表している。なお、読み上げの際には、複数の音声を同時に再生するのではなく、SNS ストリームを構成する投稿を順番に 1 つずつ再生する。

4.2 ユーザーのグループ化

本節では、対象とするユーザーのタイムラインを構成するツイートの投稿者を自動的にグループ化する方法を説明する。Twitter では、ユーザーは友人や知人、興味のあるユーザーをフォローすることで投稿を閲覧する。本論文で提案する手法では、フォロー関係を利用してユーザーをグループ化する。

グループ化は、グラフクラスタリングを用いたコミュニティ抽出によって行う。そのために、ノードが「ユーザー」、エッジが「フォロー」を表すグラフを考える。グラフクラ

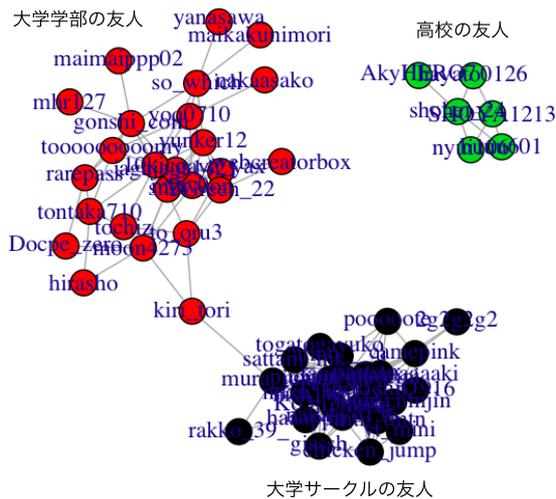


図 6 グループ分けの結果の例

Fig. 6 Examples of clusters on the followees of a user.

スタリングには、いくつかの手法が提案されているが、本研究では、貪欲法を用いたアルゴリズムを利用する [11]. この手法は、クラスタリングの精度を測る指標として、モジュラリティ Q を使い、この値が高くなるように隣接ノードのペアをまとめていく手法である。モジュラリティ Q は式 (1) で定義される。 e_{ii} は「総エッジ本数」に対する、コミュニティ i 内部における「ノードごとのエッジ本数の総和」を表し、 a_i は「総エッジ本数」に対する、コミュニティ i から他のコミュニティに向けて張られているエッジ本数の割合を表す。

$$Q = \sum_{i=0} (e_{ii} - a_i^2) \quad (1)$$

図 6 は、あるユーザがフォローするユーザをクラスタリングした結果の例である。ノードは「ユーザ」、エッジは「フォロー」を表している。赤、黒、緑の色は、グループを表し、この例では 3 つのグループに分かれていることが分かる。なお、赤は大学学部の友人、黒は大学サークルの友人、緑は高校の友人のグループを表している。

5. プロトタイプシステム

我々は、4 章で述べた手法に基づいて、プロトタイプシステム Audio Twipp を実装した。図 7 にシステムの概要を示す。ユーザは、Twitter でアカウント認証を行い、システムは、フォロワーデータを自動的に収集し、ユーザがフォローする他のユーザ (フォロワー) をデータベースに登録する。次に、データベースに登録されたフォロワーに対して、グループ化を行い結果をデータベースに登録する。読み上げ時には、システムは Twitter からタイムライン (SNS ストリーム) の取得を行い、投稿者のグループごとに異なる定位から投稿を読み上げる。読み上げる内容はユーザ名と投稿内容である。

Audio Twipp は iPhone アプリケーションであり、

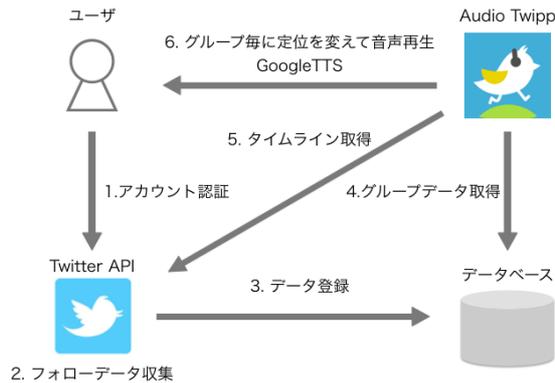


図 7 システム概要図

Fig. 7 An overview of the proposal system.

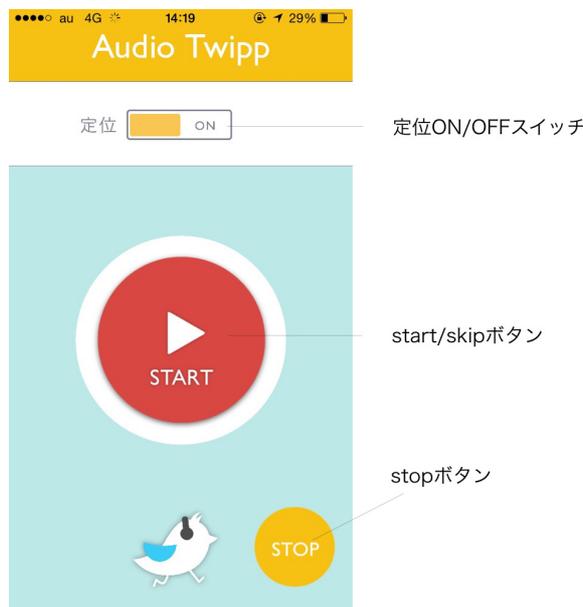


図 8 インタフェース画面

Fig. 8 A snapshot of the interface.

Objective-C 言語で実装を行った。テキストの読み上げには、オープンソースの Google TTS (Text To Speech) API を用いた。図 8 は Audio Twipp のユーザインタフェース画面である。ユーザは画面上部のスイッチによって、定位の ON/OFF を切り替えることができる。中央には、start/skip ボタンがあり、それを押すことによってユーザは興味のない投稿が流れたときに聴き飛ばすことができる。右下には stop ボタンを配置し、ユーザはこのボタンで再生を停止することができる。

6. 評価実験

提案手法の有効性を検証するために、被験者実験に基づく評価を行った。評価の内容は以下の 3 事項であり、それぞれの事項に対して個別の実験を行い評価を行った。

- (1) フォロー関係を利用したグループ化の有効性の評価
- (2) 音声読み上げ時の定位の個数に関する評価
- (3) 定位を利用した音声ブラウジング手法の有効性の評価

6.1 フォロー関係を利用したグループ化の有効性の評価

提案手法で行うフォロー関係を利用した投稿者のグループ化が適切であるかについて評価する。

6.1.1 実験手法

被験者それぞれに対して、各自の Twitter アカウントに対して、提案手法によりフォローしている他ユーザのグループ化を行った。そして、被験者に、グループ分けした結果を見せ、それぞれのグループに名前をつけるように依頼した後、それぞれのグループごとに、名前をつけたグループ名に適さないユーザを列挙するように依頼した。被験者は、毎日 Twitter を利用している 20 歳から 25 歳までの 10 人 (男 7 人, 女 3 人) である。

6.1.2 実験結果と考察

図 9 の棒グラフは抽出されたグループ数の平均を表しており、エラーバーは標準偏差を表している。抽出されたグループ数の平均は 4.2 であり、標準偏差は 0.63 であった。

各グループの構成要素数に対する、正しくグループ化されたと考えられる構成要素数の割合をグループ化の精度と定義する。実験で得られた各グループに対して、グループ化の精度を計算した。図 10 の棒グラフはグループ化の精度の平均値を表しており、エラーバーは標準偏差を表している。平均の精度は 0.81 であり、標準偏差は 0.09 であった。

上記の結果から、提案手法で得られたユーザグループはユーザにとってクラスタの意味が理解でき、約 80% の精度でユーザをグループ化できることが分かった。

6.2 音声読み上げ時の定位の個数に関する評価

提案手法では、グループごとに読み上げ時の音像の定位を変化させるが、グループの数が増大した場合には、グループを区別できないことや、誤って判断してしまう可能性がある。そこで、定位の数を、3 個、4 個、5 個とした場合の

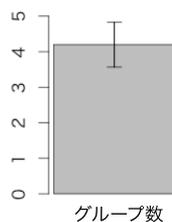


図 9 抽出されたグループ数の平均値

Fig. 9 The average number of groups extracted.

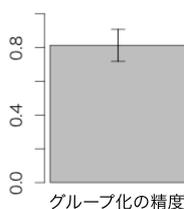


図 10 グループ化の精度の平均値

Fig. 10 The average precision of grouping.

ユーザの認識精度の違いを被験者実験により評価した。

6.2.1 実験手法

本実験では、音源の個数が 3, 4, 5 の場合を対象として実験を行った。実験の音声データには、Twitter から取得したツイートを利用した。図 11 にそれぞれの場合の音源の定位を示す。図 11 では、音波記号で表しているが、ツイートが読み上げられる定位を表している。被験者は、それぞれの状況において練習を行い、その後ランダムに定位する 30 個の音源を聴取した。そして、被験者には、聴こえたと感じた音源の定位 1 つ 1 つを回答してもらった。被験者は、21 歳から 26 歳までの男女 10 人 (男 7 人, 女 3 人) である。使用したスマートフォンは、Apple 社の iPhone5s であり、イヤホンは、iPhone 付属のイヤホンを用いた。

6.2.2 実験結果

図 12 に実験結果を示す。棒グラフは、それぞれの定位における平均正答率を表し、エラーバーは、標準偏差を表している。定位が 3 つの場合の平均正答率は 1.00、定位が

定位が 3 つの場合



定位が 4 つの場合



定位が 5 つの場合



図 11 定位ごとの概念図

Fig. 11 The directions of sounds for three types of experiments.

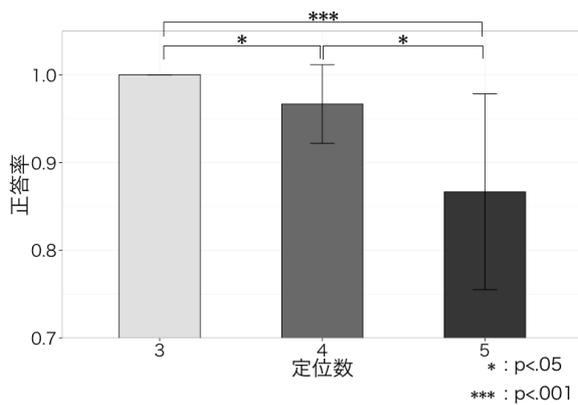


図 12 定位ごとの正答率

Fig. 12 The precisions of three types of experiments.

4つの場合は0.97、定位が5つの場合は0.87であった。それぞれの定位に対するデータの分布が正規分布に従うかどうかをコルモゴロフ-スミルノフ検定を用いて検定を行った結果、定位が3の場合には、 $p = 4.11 \times 10^{-9}$ となり、正規分布に従うと判断することはできなかった。したがって、ノンパラメトリックなマンホイットニーのU検定を用いて、有意差があるかの検定を行った。定位が3つの場合と4つの場合、定位が4つの場合と5つの場合、定位が3つの場合と5つの場合の間には、それぞれ、有意水準5% ($p = 0.035$)、5% ($p = 0.024$)、0.1% ($p = 0.00022$)で有意差がみられた。

6.2.3 考察

実験結果から、定位が3つのときは、ユーザは音源定位を間違えることなく回答することができるということが分かる。

被験者に対して、3種類の定位での読み上げを体験した後で、感想を自由記述で書いてもらったところ、以下のような結果が得られた。定位が4つ、5つのときに、「判断に時間がかかった」という意見や、「集中力が必要」といった意見があった。これは、正答率が下がっていることと合わせると、定位が4つ、5つにおいては、2チャンネルステレオ音響空間で利用可能にするためには改善が必要である。たとえば、立体音響空間を用いた音声提示をすることにより、改善することができると考えられる。

以上より、定位が3つの場合は、正確に定位を認識できるが、定位が増大するにつれて認識精度が低下することが分かった。

6.3 定位を利用した音声ブラウジング手法の有効性の評価

読み上げ音声をグループごとに異なる定位に対応付けて再生することの効果を検証するため、プロトタイプシステム Audio Twipp を用いて、被験者実験を行った。比較手法（ベースライン）としては、すべてのツイートの読み上げ音声を同一の定位（中央）で再生する手法を利用した。

なお、定位の数は、ユーザが正しく区別できる3定位を利用した。

6.3.1 実験手法

被験者に、次に述べるタスクを行ってもらった。まず、被験者は、練習のために、あらかじめ用意したアカウントで、定位を変化させる場合と変化させない場合で、音声でツイートを聴いてもらった。本実験では、3つのグループを右、中央、左の3つの定位に分けて提示した。

そして、本人のアカウントを用いてプロトタイプシステムで、定位を変化させる提案手法と、定位を変化させない比較手法（ベースライン）の音声ブラウジングによる聴取を1分間ずつ行った。ここで、1分間の聴取したツイート数、聴き飛ばし回数、聴き飛ばすまでの時間、聴き飛ばさなかったツイートの再生時間に関するデータを取得した。なお、ツイートの長さや内容に偏りがあるため、この試行を3回繰り返した。また、ユーザのフォローしているユーザが、4グループ以上のグループになった場合、それぞれのグループ化の精度が高い上位3つのグループを実験に用いた。

さらに、ユーザビリティを評価するために、被験者にはタスクの後、以下の7項目について5段階のアンケートに回答してもらった（5が最も良い評価を表す）。

- (1) 機械音声による読み上げで、ツイートの内容は理解できましたか。
- (2) 機械音声による読み上げで、ツイートは聞き取れましたか。
- (3) 提案手法において、音声の聴こえてくる位置が変化したことを認識できましたか。
- (4) 提案手法において、グループごとに音の聴こえてくる位置を分けて提示しましたが、グループを認識できましたか。
- (5) 提案手法において、聴こえてくる位置が変わることによって聞き取りやすくなりましたか。
- (6) 提案手法において、グループごとに音の聴こえてくる位置を分けて提示しましたが、聴き飛ばす手がかりになりましたか。
- (7) 比較手法のアプリケーションと、提案手法のアプリケーション、どちらを使いたいですか。

被験者は、20歳から25歳までの10人（男性7人、女性3人）であった。使用したスマートフォンは、Apple社のiPhone5sであり、イヤホンは、iPhone付属のイヤホンを用いた。

6.3.2 実験結果

ユーザが聴取したツイート数に関する結果を図13、聴き飛ばし回数に関する結果を図14、聴き飛ばすまでの時間に関する結果を図15に示す。エラーバーは、標準偏差を表している。

ユーザが聴取したツイート数の平均値は、提案手法で15.4ツイート、比較手法で11.4ツイートであった。提案手法と比較手法それぞれの結果に対して、コルモゴロフ-スミルノフ検定を用いて、それらが正規分布に従うかどうか

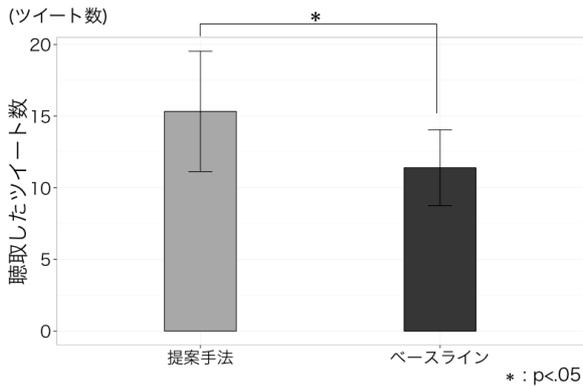


図 13 聴取したツイート数の平均値

Fig. 13 The average numbers of browsed tweets.

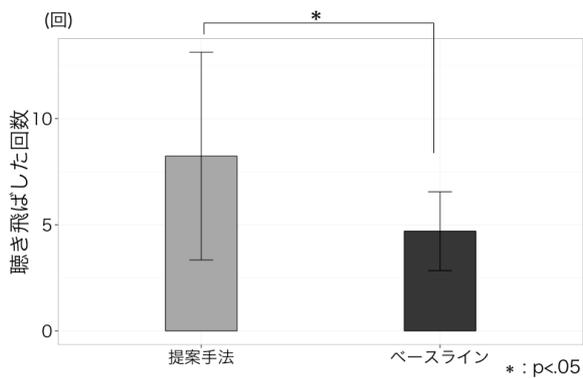


図 14 聴き飛ばした回数の平均値

Fig. 14 The average numbers of skipped tweets.

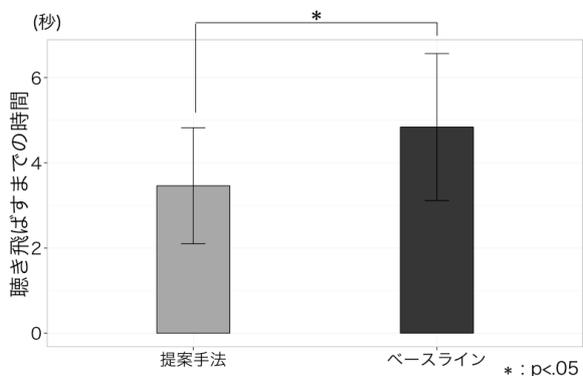


図 15 聴き飛ばすまでの時間の平均値

Fig. 15 The average times for skipping tweets.

について検定を行った結果、それぞれ、 $p = 0.93 > 0.05$ および $p = 0.96 > 0.05$ となり、ともに正規分布に従うという仮説が採択できる。そこで、提案手法と比較手法の実験結果に対して、対応のある t 検定の片側検定を行った結果、有意水準 5% ($p = 0.018$) で有意差が存在した。したがって、提案手法を利用することで、比較手法を利用したときよりも、ユーザは 1 分間により多くのツイートを聴取できるといえる。

聴き飛ばし回数の平均値は、提案手法で 8.3 回、比較手法で 4.7 回であった。提案手法と比較手法それぞれの結果

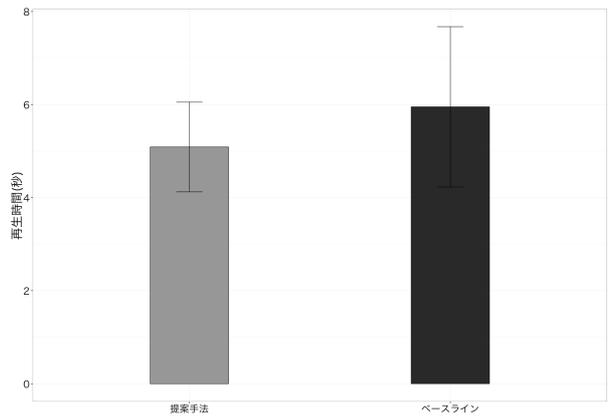


図 16 聴き飛ばさなかったツイートの再生時間の平均値

Fig. 16 The average times for playing non-skipped tweets.

に対して、コルモゴロフ-スミルノフ検定を用いて、それらが正規分布に従うかどうかについて検定を行った結果、それぞれ、 $p = 0.51 > 0.05$ および $p = 0.94 > 0.05$ となり、ともに正規分布に従うという仮説が採択できる。そこで、提案手法と比較手法の実験結果に対して、対応のある t 検定の片側検定を行った結果、有意水準 5% ($p = 0.028$) で有意差が存在した。したがって、提案手法を利用することで、比較手法を用いた場合よりも、ユーザは 1 分間により多くのツイートを聴き飛ばすことができるといえる。

聴き飛ばしを行うまでの時間の平均値は、提案手法で 3.5 秒、比較手法で 4.9 秒であった。提案手法と比較手法それぞれの結果に対して、コルモゴロフ-スミルノフ検定を用いて、それらが正規分布に従うかどうかについて検定を行った結果、それぞれ、 $p = 0.45 > 0.05$ および $p = 0.80 > 0.05$ となり、ともに正規分布に従うという仮説が採択できる。そこで、提案手法と比較手法の結果に対して、対応のある t 検定の片側検定を行った結果、有意水準 5% ($p = 0.014$) で有意差が存在した。したがって、提案手法を利用することで、比較手法を利用した場合よりも、ユーザが聴き飛ばしを行うまでの平均時間が短くなるといえる。

また、提案手法と比較手法において、聴き飛ばさなかったツイートの再生時間の平均値を計算した結果を図 16 に示す。聴き飛ばしを行わなかったツイートの再生時間の平均値は、提案手法で 5.1 秒、比較手法で 6.0 秒であった。提案手法と比較手法それぞれの結果に対して、コルモゴロフ-スミルノフ検定を用いて、それらが正規分布に従うかどうかについて検定を行った結果、それぞれ、 $p = 0.95 > 0.05$ および $p = 0.68 > 0.05$ となり、ともに正規分布に従うという仮説を採択できる。そこで、提案手法と比較手法の結果に対して、対応のある t 検定の両側検定を行った結果、 $p = 0.17 > 0.05$ となり、有意水準 5% では有意差があるとはいえなかった。したがって、提案手法を用いた場合と、比較手法を用いた場合において、「聴き飛ばし」を行わなかったツイートの再生時間に違いがあることは示されなかった。

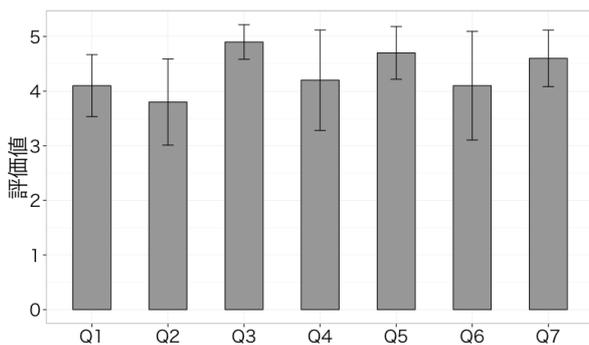


図 17 アンケート結果

Fig. 17 The results of questionnaires.

図 17 にアンケートの結果を示す。すべての項目において、中間値 3 以上の良好な評価を得た。項目 1, 2 の結果から、機械音声によるツイートの読み上げでも内容の理解をすることが可能であることが分かる。項目 3 では、イヤホン環境下においても、音声の定位を認識することが可能であることが分かる。項目 4, 5, 6 からは、提案手法において、定位を変化させることが聴き飛ばす行為を行う要因となっているという評価が得られた。

6.3.3 考察

上記の聴取したツイート数、聴き飛ばす回数、聴き飛ばすまでの時間のそれぞれで比較手法と提案手法の間に有意差が存在しており、被験者の提案手法に対する主観評価の値も良好であった。さらに、聴き飛ばさなかったツイートの再生時間に関しては有意差があるとはいえなかった。したがって、提案手法によって、聴き飛ばす回数が増え、聴き飛ばすまでの再生時間を短縮可能となり、結果的に多くのツイートを聴取可能となったことから、提案手法は SNS ストリームを対象とした音声ブラウジングの効率化に効果があると考えられる。

アンケートの自由記述からいくつか知見を得ることができた。その中に、「音声を聴き取りにくいことがあった」という記述があった。聴き取りやすさには個人差があるが、より精度の高い読み上げが必要である。「内容を聴いて興味があるかどうか判断した」という記述もみられた。今回は、人の関係を定位で表現したが、取得したツイートデータを解析し、トピックごとにグループ分けをする手法も可能であると考えられる。

今後、トピックによるグルーピングの手法についても検証する必要がある。本手法では、3つの定位を用い、3つのグループでの実験を行った。しかし、実際は4つ以上のグループになることもあるため、定位の数を増やして実験を行っていく必要がある。

7. おわりに

本論文では、効率的な SNS ストリームの「聴き飛ばし」手法を提案し、被験者実験により有効性を評価した。我々

は、まず、視覚ブラウジングにおける「読み飛ばし」の分析を行い、ツイートを構成するユーザアイコンに注目し、『タイムライン内のツイートを「読み飛ばし」をするかを判断する際に、ユーザアイコンが重要な手がかりとなっている』という仮説を立て、被験者実験によりこの仮説が正しいことを明らかにした。そして、この結果から、Twitter のフォロー関係を分析しユーザにとって意味のあるグループに分け、グループごとに音源の定位を変化させて提示することで、より短時間に「聴き飛ばし」を行うことを可能にする手法を提案した。この手法では、ユーザのフォロー関係に基づいて、ユーザをグループ化し、ユーザのグループに応じて異なる定位において読み上げ音を再生することにより、ユーザが効率的に、ユーザにとってのツイートの価値を推定し、「読み飛ばし」を行えるかどうかを決定することを支援する。

提案手法の有効性を評価するために、iPhone 上で動作するプロトタイプシステムを実装し、被験者実験を行った。被験者実験により、本論文で提案した定位を利用した音声ブラウジング手法は、聴取したツイート数、聴き飛ばし数、聴き飛ばしまでの時間において、5%の棄却率で有意差がみられた。さらに、アンケート結果において、定位を変化させることで、グループごとの情報が伝達され、それにより聴き飛ばすきっかけになるという評価を得た。これらの結果から、グループごとに定位を変化させる手法が効率的な SNS ストリームの音声ブラウジングに有効であることが分かった。

今後の課題として、まず、グループ数が増加した場合の音声提示手法について考慮する必要がある。今回の音声読み上げ時の定位の個数に関する評価実験で、「左右の定位は理解しやすいが、中央付近の音源定位は、隣接する定位の後に提示されたときには分かるが、離れた定位から提示されたとき、判断しにくい」という意見があった。この課題の解決方法として、中央にガイド音を鳴らし基準を設ける手法や立体音響空間を用いた手法が考えられる。関連研究では、立体音響空間を用いた研究も多くあり、2チャンネルステレオ音響空間では判断しにくい定位も区別することができるのではないかと考える。

また、ツイートの内容に基づいたグループによる音声提示を検討したいと考えている。今回の評価実験における自由記述のアンケート結果に「内容を聴いて興味があるかどうか判断した」というものがあった。本研究では、ツイートの内容は考慮していないが、ツイートをあらかじめ分析し、内容に基づいたグループ化を行い、定位を変えて提示することにより、効果的な音声ブラウジングが実現できる可能性がある。

グループごとに「聴き飛ばし」の発生率に偏りがある場合、それぞれのグループからの投稿に対して発生率に基づいて優先度を設定し、優先度の高い投稿を先に読み上げた

り、優先度の低い投稿は読み上げを見送ったりする等の機能により、さらに効率的な音声ブラウジングが実現できる可能性がある。今後、被験者実験により、グループごとに「聴き飛ばし」の発生率に偏りがあるかを調査し、こうした手法についても検討を行っていく予定である。

本手法は、歩行中、運転中、家事の最中等での利用を想定しているが、それらの状況で使用するためには、今回のプロトタイプシステムで採用したような、ボタンによる「聴き飛ばし」の操作は現実的ではない。近年、Apple Watch や Google Glass 等のウェアラブル端末が登場している。本研究では、スマートフォン上での実装だったが、将来的にウェアラブル端末での発展も考えることができる。上野ら [12] は、スマートフォンを用いた音声 Twitter システム VoiTwi の試作を行っており、ここでは、スマートフォンを振るジェスチャを使ってリストを変化させる手法を採用している。このように、聴き飛ばしボタンだけでなく、傾きやスマートフォンの傾き等のユーザの自然な振舞いによって聴き飛ばしを行う機能についても検討する予定である。

謝辞 本論文を執筆するにあたり、本論文を大変丁寧に査読していただき、大変有用なコメントをいただいた査読者の方々に感謝いたします。また、本研究を進めるにあたり、活発に議論していただいた、九州大学牛尼研究室の皆様にも感謝いたします。

参考文献

- [1] 土岐真里奈, 牛尼剛聡: ソーシャルストリーム閲覧時の振舞いを利用したユーザプロフィール構成手法, 情報処理学会論文誌データベース, Vol.6, No.4, pp.35-45 (2013).
- [2] 海老名名毅, 猪木誠二, 河合隆芳: 聴覚フィードバックを用いた GUI オブジェクトの検索について, 電子情報通信学会技術研究報告, 95, pp.61-66 (1996).
- [3] Brewster, S., Bell, M., Hall, M. and Tasker, S.: Multimodal 'eyes-free' interaction techniques for wearable devices, *Proc. CHI '03*, pp.473-480 (2003).
- [4] Sawhney, N. and Schmandt, C.: Design of Spatialized Audio in Nomadic Environments, *Proc. International Conference on Auditory Display (ICAD)*, pp.109-113 (1997).
- [5] Goose, S. and Moller, C.: A 3D only interactive Web browser: Using spatialization to convey hypermedia document structure, *Proc. ACM Multimedia '99*, pp.363-371 (1999).
- [6] 河本敏志, 竹村治雄, 片山善章, 萩原兼一, 横矢直和: 音声メニュー同時提示方法の提案と評価, 情報処理学会研究報告, HI57-59 (1995).
- [7] Schmandt, C. and Mullins, A.: AudioStreamer: Exploiting Simultaneity for Listening, *Proc. CHI '95*, pp.218-219 (1995).
- [8] 小林 稔, クリス シュマント: 時間-空間マッピングによる音声ブラウジングツール, 情報処理学会論文誌, Vol.39, No.5, pp.1285-1296 (1998).
- [9] 飯田一博, 森本政行 (編著), 日本音響学会 (編): 空間音響学, コロナ社 (2010).
- [10] 岩永章吾, 牛尼剛聡: 立体音響空間上でのカクテルパーティ効果を利用したソーシャルストリームの効率的なブラウジング手法, DEIM 2013 論文集, B6-6 (2013).

- [11] Clauset, A. et al.: Finding community structure in very large networks, *Phys. Rev.*, E70, 066111 (2004).
- [12] 上野大樹, 安村通見: VoiTwi: スマートフォンを用いた音声 Twitter システムの試作と研究, 電子情報通信学会技術研究報告 WIT, 福祉情報工学, Vol.110, No.384, pp.1-6 (2011).



岩永 章吾

1989 年生。2013 年九州大学芸術工学部芸術情報設計学科卒業。2015 年同大学大学院芸術工学府芸術工学専攻修士課程修了。同年凸版印刷 (株) 入社。



牛尼 剛聡 (正会員)

1970 年生。1999 年名古屋大学大学院工学研究科情報工学専攻博士課程後期課程満了。1999 年九州芸術工科大学芸術工学部助手。2011 年九州大学大学院芸術工学研究院准教授。日本データベース学会, 電子情報通信学

会, ヒューマンインタフェース学会, ACM, IEEE-CS 等各会員。



角谷 和俊 (正会員)

1988 年神戸大学大学院工学研究科修士課程修了。同年松下電器産業 (株) 入社。ソフトウェア開発環境, マルチメディアデータベース, データ放送の研究開発に従事。1998 年神戸大学大学院自然科学研究科博士後期課程 (情報メディア科学専攻) 修了。1999 年神戸大学都市安全研究センター都市情報システム研究分野講師, 2000 年同助教授。2001 年京都大学大学院情報学研究科社会情報学専攻助教授。2004 年兵庫県立大学環境人間学部環境人間学科教授。2015 年関西学院大学総合政策学部メディア情報学科教授, 現在に至る。博士 (工学)。IEEE Computer Society, ACM, 電子情報通信学会, 日本データベース学会等各会員。

(担当編集委員 土方 嘉徳)