

実況 lizer: スポーツ実況のメタファに基づいた ライブ都市センサデータの可聴化手法の提案

井上 義之^{1,a)} 米澤 拓郎^{2,b)} 大越 匡^{2,c)} 中澤 仁^{3,d)}

概要: 本稿は、都市センサデータをスポーツ実況のメタファに基づき可聴化し、ユーザに伝達する実況 lizer を提案する。現在 IoT の発達やスマート端末等の普及により、多種多様かつ大量の都市データが即時に収集可能となった。都市データをユーザに伝達する際には、データの可視化手法が多く提案・利用されているが、可視化はユーザの視覚を専有してしまうため、他業務を遂行しながら情報を伝達することが難しいという問題がある。本研究では、ユーザの認知負荷をおさえつつ、重要な情報を伝達することが可能な、都市データの可聴化手法を構築する。同手法は、行政職員など都市データを必要とするユーザに対し、TV やラジオで広く親しまれているスポーツ実況のメタファに基づいた効果的な情報伝達を可能とする。

Jikkyolizer: Auditory Display for Live City Sensor Data Based on Metaphor of Sports Telecast

YOSHIYUKI INOUE^{1,a)} TAKURO YONEZAWA^{2,b)} TADASHI OKOSHI^{2,c)} JIN NAKAZAWA^{3,d)}

1. はじめに

近年、知的都市マネジメントを実現するため、都市中に設置されたセンサデバイスなどから多種多様なライブデータが収集可能となってきた。例えば、都市を走る車両が取得した位置情報や環境情報、住民や行政職員が投稿したレポート情報 [1] に加え、Web から抜き出した都市に関する情報 [2] などが都市ライブデータにあたる。これらの都市ライブデータの収集、分析を行い、その結果を自治体職員や住民、周辺企業など多様なユーザに伝達することにより、効率的な都市マネジメントやサービスの提供が期待されている。例えば、平常時には都市内の環境状況や混雑

状況をユーザに提供することにより、安心安全を高めるナビゲーションサービスや、住民への都市イベントレコメンドサービスが構築されたり、災害時には行政職員の災害対策に対する意思決定の支援や住民に対する行動の指針を与えることができる。本研究では、これらの都市データを、いかにユーザに伝達するか、に関して取り組む。

収集されたデータはそのまま、もしくは分析・加工された形でユーザに提供されるが、現在は主に可視化による情報提示がなされている。例えば、London Dashboard [3] ではテキストやグラフを用いたダッシュボード形式の可視化がなされており、[5] では地図上にデータをマッピングする可視化がなされている。視覚はヒトの五感の中で最も大きな割合で情報を受け入れていると言われており、多くの情報を効率よく伝達するためには、可視化による情報提示を行うことは重要である。一方、可視化された情報を受け取るためには、視覚を専有されてしまうことになるため、他の手段での情報提示手法も今後より重要となる。特に本研究では行政職員を対象とし、可視化以外での情報提示手法に関して議論を行う。様々な都市マネジメントに従事する行政職員に対して都市のライブデータを効果的に提供す

¹ 慶應義塾大学 総合政策学部
Faculty of Policy Management, Keio University
² 慶應義塾大学大学院 政策・メディア研究科
Graduate School of Media and Governance, Keio University
³ 慶應義塾大学 環境情報学部
Faculty of Environment and Information Studies, Keio University
a) tigerman@ht.sfc.keio.ac.jp
b) takuro@ht.sfc.keio.ac.jp
c) slash@ht.sfc.keio.ac.jp
d) jin@ht.sfc.keio.ac.jp

るためには、彼らの普段の業務遂行を阻害することなく、有益な情報を提供する必要がある。そのためには、視覚を専有することなく、一方で彼らに気づきを与えることが可能な情報提示手法が求められる。このような状況下において、視覚以外に使用できる感覚は、聴覚・嗅覚・味覚の3つとなる。本研究では、この中でも特に聴覚に訴えることに着目した。その理由として、視覚の次に情報を受け取る割合が多いのが聴覚であること、また言葉を利用することでより多くの情報量を形式的に伝えることができる、ということが挙げられる。

本研究では、行政職員の認知負荷を抑えつつ都市に関する情報を伝達する手法として、都市ライブデータの効果的な可聴化手法を構築することを目的とする。特に本稿では行政職員を対象とし、行政職員が業務遂行中に情報を伝達可能とするため、彼らの「認知負荷を抑えつつ」「重要な情報を伝達する」という2点を満たす手法を構築する。この目的を達成するため、本研究では都市ライブデータの可聴化手法として、スポーツ実況のメタファに基づく可聴化手法を提案する。スポーツ実況の要素を取り入れる理由は主に以下の二つである。第一に、スポーツ実況はこれからどう動くかわからないライブデータを正確に伝える方法であり、第二に、ライブデータを過去情報と関連させ、ライブデータに意味を付与させた後に視聴者へ伝える方法が含まれているからである。本研究では、スポーツ実況のメタファを取り入れた可聴化手法が効果的であるか検証するため、可聴化システム実況 lizer を提案する。本稿では、まずスポーツ実況のメタファについて分析を行い、次にその分析に基づき提案する可聴化システム、実況 lizer の概要について述べる。また、実況 lizer を都市ライブデータ共有基盤である SOXFire[4] を用いて実装を行い、実況 lizer の効果をユーザ実験により評価する。結果、ユーザの置かれた状況に応じ、実況 lizer が有効に作用することがわかった。本研究の貢献は、以下の3点である。

- スポーツ実況のメタファに関して整理し、都市ライブデータに適用可能な要素を考察したこと
- 都市ライブデータを実況形式で可聴化するシステム、実況 lizer を構築したこと
- ユーザ実験により、実況 lizer の有効性を評価したこと

2. スポーツ実況のメタファに基づく都市ライブデータ可聴化システムの提案

本章では、本稿で提案する都市ライブデータ実況システム、実況 lizer について説明を行う。まず、本研究が対象とするスポーツ実況のメタファに関して整理、考察を行なったあと、実況 lizer で採用するスポーツ実況のメタファに関して説明を行う。

2.1 スポーツ実況のメタファ

スポーツ実況にはライブデータを正確に、過去情報と絡めて伝える方法論が含まれている。例えば、競馬中継のメディアであるフジテレビは以下のような記事を公開している。

“(競馬実況には)「騎手の名前」や「白い帽子などの視覚的情報」「途中のレースタイム」「会場の様子」そして競馬ならではのフレーズでドラマを「うたいあげる」など10項目の情報が入っていることが分かった。では、競馬ファンはどんな情報が重要と考えているのか、(中略)ファンは実況に含まれていた10項目すべての情報を求めている。”*1

上記のように、スポーツ中継の視聴者が求めている情報を伝えるためには、スポーツ実況独特な方法論が存在しており、本研究が対象とするライブデータの可聴化の際にも参考となると考えられる。本節では、実況の要素を、(1)実況コンテンツの決定、(2)実況文章の作成、(3)実況文章の音声化、の3段階に分け、スポーツ視聴を支える実況技術について考察する。

2.1.1 実況コンテンツ決定

実況を行うためには、まずはじめに、様々な事象が起きている中から実況する内容を正しく抽出・整理して実況の内容や順番を決めることが必要である。

● 情報抽出

実況者は目の前でリアルタイムに起こっている事象を一瞬の間に抽象化し、それを音声として発している。この際に行われていることは「a. 実況, b. 分析, c. 過去の情報を伝える, d. 予測, e. 会話」[7]の5つであるとされており、それぞれ、

- a. 実況：試合が動いている際のプレーの描写
- b. 分析：試合展開やプレーの分析
- c. 過去の情報を伝える：過去のデータや資料を紹介
- d. 予測：これから先の戦術や選手交代があるかを予測

e. 会話：解説者やリポーターとのやりとりを示している[6]。また、この分析や予測の際に、“実況アナウンサーは「筋書きのないドラマ」の「筋書き」をあえて描き、視聴者に勝ち負けにつながる「物語」を提供している”[6]という手法が使われている。

● 情報整理

実況者の考えによって、上述した a. e. の情報が実況される優先順位は異なる。

“冷静に状況を分析・解説していく「理論」派の実況もあれば、感情を前面に押し出して盛り上げていく「熱血」派の実況もあります。”*2

*1 競馬実況ってなんで早口なの?
<http://blog.fujitv.co.jp/newhiho/E20140531003.html>

*2 サッカー実況、好きなのは「理論」派? 「熱血」派? http://number.bunshun.jp/articles/-/81799/feedbacks?per_page=10

この二つには、実況するコンテンツの違いが存在する。「理論」派は現在の様子や、過去との比較に基づいた試合の分析などがコンテンツとなる。それに対し、「熱血」派は視聴者の感情を高ぶらせる、あるいは視聴者の心境を代弁するような言葉がコンテンツとなる。実況者間の差もあるが、競馬のような賭けている人がいるスポーツ、あるいは視聴者がそのスポーツに関する玄人が多い場合は、「理論」派で実況されることが多い。物語性よりも自分の賭けた馬の行方の方が大事であり、物語を教えてもらわなくとも目の前の情報から各々で物語を描けるためである。一方、プロレスなどのショーの側面が強い競技・視聴者に素人が多い場合は「熱血」派で実況されることが多い。

2.1.2 実況文章生成

実況するコンテンツが決まった後は、その内容を文章化する必要がある。

- 文章テンプレートへの当てはめ

実況者はある程度、文章をテンプレート化することによって、実況を形式的に伝えることが多い。音のテンプレート化には二つの種類がある。一つは情報と情報を区切るもの、もう一つは「文章の容器」を用意し、そこに実際の状況を入れていくものだ。例えば、2016年12月31日におけるボートレース多摩川の実況においては「スタートは正常、1周2マーク！」というセリフが12レース中11レースに出現している。これは今まで競艇において直線の攻防を実況していたが、これからはターンする時の部分を実況すると実況者が合図を出し、情報と情報を区切った例である。そして2017年7月18日のボートレース常滑では「インコースから1号艇の(選手名)先行し、2コースから2号艇の(選手名)は引いた。(真ん中・外)を攻めていくのは(選手名)バックへ入った！」というテンプレートが12レース中5レースにおいて使われていた。変わったのは(括弧)でくくった部分のみで、他は殆どそのままの文章で実況されていた。これは長い「文章の容器」に対して、目の前の状況を埋めて実況を行い、情報を形式的に伝達した例である。

- 状況をまとめる言葉の使用

専門用語の多用がスポーツ実況には多く見られる[6]。これは多くの情報の取捨選択が求められる状況において、(1) 尺を取らない言葉(=専門用語)で表現することが重要であること、および、(2) 専門用語の方が、スポーツに詳しい人間により直感的に伝えられること、という二つの理由がある。時間の限られた実況において尺を取らないということは大事で、「聞きやすさが求められるニュースと比較してみると、競馬実況は1分間で577文字なのに対し、ニュースは358文字。競馬実況はニュースの約1.6倍だった」*3との記事もある。早口すぎると情報が正しく伝わら

ない可能性もあるため、情報をまとめた専門用語は多用せざるを得ない。また、専門用語は事象をわかりやすく伝えるという側面も持っている。サッカーの実況においてゴール前、「クロスがきた！」といったとき、クロスの意味を知る視聴者はピッチの左右端から内側へボールが飛んできた理解する。「クロスがきた！」という実況を別の短い文章で「左端からボールが飛んできた！」と実況すると、それは画面の左側からボールが来たのか、注目している選手の左側からボールがきたのかわからない。一方、「ピッチの左側からゴール前へのボールがきた！」とより長い文章に実況しても、おそらくそれを全て聞き取って情報を組み立て直すことは難しいだろう。実況に耳を済ましていれば可能かもしれないが、試合を見ながら長い文章を認知することは厳しい。つまり、専門用語はより直感的に情報を伝えることに成功していると言える。

2.1.3 実況文章の音声化

実況文章は最終的に音声化されて視聴者に届けられる。この際、伝える文章の内容に加え、より視聴者にその内容が伝達されるよう、文章の表現方法が工夫されている。

- 声の高さ

実況者は珍しいものを、驚きをもって表現する場合がある。人は珍しいものを見たとき、驚く。これを逆手にとつて、珍しい状況になったとき驚きをもって表現し、伝えている状況が珍しいことを示している。そしてこの驚きは、声の高さを変えることによって表される場合が多い。“感情の高ぶりはピッチとテンポで表現できる”*4と、コミュニケーションアナリストも分析している。実況者も同様に、声の高さを普段と変えることで感情の高ぶりを表現する手法を用いている。珍しいことが起こり声の高さを変える時、実況者は高い声や低い声のどちらかばかりを使いすぎないようにし、視聴者を飽きさせないようにしている。

- 声の大きさ

試合において、常に視聴者に多くの注意を向かせることは、視聴者を疲れさせてしまう。大事な場面だけ声を大きくしメリハリをつけることで、特に視聴者の注意を引かせることは常套手段である。また、スポーツにおいて大事な場面というものは、多くの場合興奮を伴うものである。大事な場面で声を大きくし、注意を引くとともに視聴者に興奮をもたらす作用があると考えられる。声の大きさについて、実況者によって大きく2つの種類に分けられる。上述した情報整理でも述べたが、実況には「理論」派と「熱血」派が存在する。「理論」派の実況者はあまり声の大きさは変えない。逆に「熱血」派は重要な場面で声の大きさを大きさに変える。そして“中継番組のニーズに合わせ、アナウンサーは実況の仕方を変えます。”*5と実況者が話して

*3 競馬実況ってなんで早口なの?
<http://blog.fujitv.co.jp/newhiho/E20140531003.html>

*4 上野 陽子, “第一印象は声が4割! 効果絶大な「声」の6技術”
<http://president.jp/articles/-/13681?page=2>

*5 スポーツ実況の絶叫連呼 事情があつて

いるように、その競技の玄人が見る番組は「理論」派な実況が好まれ、素人も見る番組は「熱血」派な実況が好まれる傾向にある。

- 声の速さ

実況者は、上述した文章のテンプレート化にそって情報を伝える際、声の速さを変化させる場合がある。テンプレートにおける「文章の容器」はただ情報の構造を示しているのみで、情報そのものを示してはいない。したがって、「文章の容器」そのものの部分は速く喋り、「文章の容器」に入れられた中身の部分をゆっくり喋ることで、情報の内容を際立たせて実況することができる。また、テンプレート化されていない文章についてはいずれも重要なので、速さを抑えて実況を行う。上述した「状況をまとめる言葉」が実況では多く使用され、状況を伝える尺を確保するという技が使われている状況を述べたが、これと基本的な考え方は同じである。

2.2 実況 lizer におけるスポーツ実況のメタファの適用

上述したスポーツ実況のメタファに基づき、都市に関するライブデータを実況する可聴化システム実況 lizer を本研究では提案する。実況 lizer が想定する都市データとして対象とするのは、環境データ（天気、温度、湿度）、人の流れに関するデータ（店舗の待ち時間、駐車場の空車状況）、産業データ（家賃相場、ダムの残水量）などが含まれる。また、実況される目的は、都市のライブデータを、他の作業をしながら実況 lizer を聞いている市の行政職員に伝えることである。以下、実況の各要素の、実況 lizer における適用方法について述べる。

2.2.1 実況コンテンツ決定

- 情報抽出

2.1.1 における、a 実況・b 分析を実況 lizer で扱うコンテンツとする。なお今後、a 実況は本論文では曖昧さ回避のため、「生実況」と言い換える。「b 分析」は「状況分析」と言い換える。生実況は得られた情報をそのまま伝える実況を意味し、状況分析は現在の情報が持つ意味を噛み砕いて伝える実況を意味する。例えば、「藤沢市の家賃相場」の値が「11.11」万円であったとして、そのデータをそのまま伝えるだけでは、そのデータの解釈は難しい。家賃相場は上がっているのか、他の市と比べてどうなのか、などが分かれば「11.11」万円がどういう値なのかがわかる。c 過去の情報を伝える目的は、過去と比較して現在はどうか？という情報を視聴者に伝達することであり、状況分析の形態を含め視聴者に伝えた方が、より高度な理解を促す情報伝達となる。また、d 予測に関しても、その情報に基づいた将来の予測や関連情報（例：明日の天気）などの情報を付加して実況に加えることで、現在のデータの解釈に対する

示唆を与えることができる。よって、これらの要素は、実況 lizer にとって重要な要素であると考えられる。一方、本稿の実況においては詳細な過去・将来予測に関してはより付加的な情報と捉え、まずは生実況と単純な状況実況の実現を対象とする。e 会話であるが、スポーツ実況においては実況者とコメンテーター（専門家）という役割を置く場合が多く、この会話がスポーツの描写の印象を高める役割を果たしている。都市ライブデータにおいてはその分析などもシステム内部で行えるため、実況者とコメンテーターという役割分担を擬似的に行うことが可能であると考えられる。一方、これを達成するためにはそもそもの分析が必要があるため、本稿ではまず実況者 1 人が存在する、という想定をおこなう。

- 情報整理

本稿で提案する実況 lizer では、「理論」派の立場から、実況を行う。本研究の実況の視聴者は市職員であり、都市情報については精通している場合が多く、ライブデータをショーの内容としてよりも実践的な情報として受け取る場合が多いと考えられるためである。一方で、生実況のみの情報と、生実況と状況分析含んだ情報とを比較した場合では、状況分析の情報が含まれている情報の方がより価値があると考えられる。生実況は単なる事実のみであり、状況分析はそれに対する解釈も盛り込まれているためである。よって、状況分析の情報を含んだ実況を、より優先的に行う方針をとる。

2.2.2 実況文章生成

- 文章テンプレートへの当てはめ

情報と情報を区切る方法は、話題があるライブデータから違うライブデータへ移ったことをユーザに示すのに適切であると考えられるので、実況 lizer へも適用を行う。また、「文章の容器」に実際の状況を入れていく方法も、ライブデータは構造化されたライブデータを運んでくるため有用であると考えられ、適用を行う。

- 状況をまとめる言葉の使用

都市ライブデータもスポーツの状況と同じく、もしくはそれ以上の多数の情報が運ばれてくるため、状況をまとめる言葉を使用する必要がある。

2.2.3 実況文章生成

- 声の高さ

スポーツ実況同様、驚きの度合いが高いものほど可聴化された声の高さを変える仕組みを用いる。本稿では、平常時より値が大きい時は高い声で、低い声の時は低い声で音声化を行う。飽きさせない工夫よりも分かりやすさを優先させることで、今回はより直感的な情報の伝達を目的とし、「高い」あるいは「低い」という声が聞き取れなくても内容が推察できるようにする。

- 声の大きさ

「理論」派な実況を心がけ、大きく言葉の大きさを変えるこ

やっていると現役アナ http://www.news-postseven.com/archives/20130213_171538.html

とは行わないこととする。一方で述語部分、つまり状況をまとめる言葉や分析結果の部分は重要な部分であるので、他の部分より少し声を大きくして音声化を行う。これは本稿においては、市行政に関して知識を持っている市職員が、情報を役立つものとして受け取るための実況を実現するためである。

- 声の速さ

「文章の容器」はスピードの速い音声を使い、実際のデータの部分を通常の音声スピードとする。

3. 設計・実装

本章では、実況 lizer の設計と実装について述べる。本研究では、都市ライブデータの情報源として、SOXFire[4]を用いる。SOXFire は慶應義塾大学で開発・運用が行われている都市ライブデータ収集・共有基盤技術であり、様々な都市データが IoT や市民・職員が有するスマート端末、また関連 WEB サイトを情報源として収集がなされている。よって、SOXFire 内に流通される都市ライブデータを対象として、そのデータの実況を可能とするシステムとして実況 lizer の設計・実装を行う。SOXFire は XMPP プロトコルの Publish-Subscribe 機能を用いたセンサデータ流通を実現しており、実際のセンサデータは XML 形式で配信される。SOXFire ではメタノードとデータノードという2つの PubSub ノードがペアとなった仮想センサノードでセンサノードが管理されており、前者が仮想センサノードのメタ情報を配信し、後者が実際のセンサデータを配信する。それぞれの仮想センサノードにはデバイス ID の他、1つ以上のトランスデューサと呼ばれるセンサ（例：温度センサ、湿度センサ等）が存在する*6。本章では SOXFire を利用した実況 lizer の設計についてまず述べ、その後実装について説明する。

3.1 設計

図 1 に実況 lizer のシステム構成図を記す。実況 lizer はまず SOXFire に存在するすべての仮想センサノードをサブスクライブし、情報抽出部において都市ライブセンサデータを受信するとともに、データの初期的なクレンジングを行う。また同時に、内部データベースに受信データの記録を行う。次に、情報整理部において、実況すべき情報に対して優先度を設定する。次に、テンプレート適用部において、文章のテンプレートに添って実データを挿入する。更に、データの解釈を短い言葉で行うため、文章修飾部において追加の文章を生成する。これらの処理によって生成された文章は音声生成部に渡され、声高、声量、速度を調整し、最終的に実況音声として出力される。以下に、各機能の詳細について説明を行う。

*6 詳細は <https://xmpp.org/extensions/inbox/sensors.html> を参照

3.1.1 コンテンツ決定部

- 情報抽出部

生実況は SOXFire から流れてきたライブデータをそのまま抽出し、情報整理のモジュールに渡す。状況分析は SOXFire から流れてきたライブデータとその同一トランスデューサの過去情報との比較、同一種類であるトランスデューサ間の比較、の二つを行う。過去情報との比較は、ライブデータの値が同一トランスデューサの過去情報と行う。同一情報との比較は、ライブデータの値が同一種類のトランスデューサの現在情報と行う。そして各々、上位 10%・下位 10%にある場合、あるいは最大値・最小値にある場合に分析結果を情報として情報整理のモジュールに渡す。また、例えば「35 度」など単位ごと値がライブデータとして流れてくる場合があるが、この場合にも対応できるように単位を取り外し、数値型での条件分岐を可能にする前処理を行う。

- 情報整理部

状況分析の結果を含んだ情報を優先的に実況させるため、以下の処理を行う。以下に処理の順番を示す。

- (1) 送られてきた情報が状況分析の情報を含んでいなければ、実況優先度を 1 とする。
- (2) 状況分析の情報を含んでいて、その結果が上位・下位 10%に関するものであれば実況優先度を 2 とし、最大値か最小値に関するものは実況優先度を 3 とする。
- (3) 情報が送られてきてから 5 分以上経過したものについては、実況の対象から外す。

3.1.2 実況文章生成部

- テンプレート適用部

情報と情報との区切れ目に、「ポーン」という機械音を入れ、情報の区切れ目を示す。生実況と状況分析情報との間にこの区切り音は入れないが、ある実況対象から違う実況対象に移る際、この区切り音を用いる。また、生実況の際には「(何秒、または何分)前、(ノード名)の(トランスデューサ名)の値が(受け取ったライブデータ)(になりました・を示しました).」という「文章の容器」を用いた。これは数種類ある「音声の容器」の中の一つであり、他の分析結果や専門用語を使うときもそれぞれの「音声の容器」を用意し、利用した。情報抽出部より抽出された情報を、「文章の容器」内に挿入する。一部、小数点の小数点以下が多くなってしまう場合が見られたので、その場合は有効数字を 2 桁として文章に組み込む。

- 文章修飾部

本機能では、状況をまとめる言葉の利用を実現する。例えば、「tenki.jp 延岡」(ノード名)の「temperature」(トランスデューサ名)が「35 度」(ライブデータ)を示した場合、そのトランスデューサ名のライブデータが「35 度」以上の場合、「猛暑日」と定義・登録しておく。そして、「31 秒前、tenki.jp 延岡の temperature の値が 35 度を示しまし

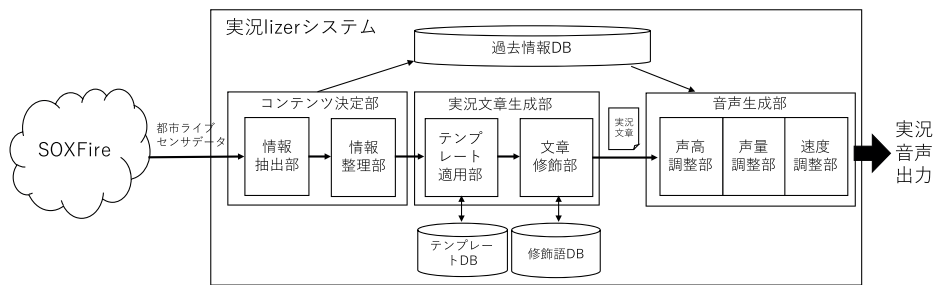


図 1 実況 lizer のシステム構成図

た. tenki.jp 延岡の temperature は真夏日です。」という文章を生成する. 事前の状況をまとめる言葉は直接データとして条件ごと登録し, データを検索し, 条件に当てはまった場合にそのまとめる言葉を文尾に付与する.

3.1.3 音声生成部

● 声高調整部

状況分析の結果より, 同一トランスデューサ内の過去情報, あるいは同一種類のトランスデューサの現在情報と比較してライブデータが上位 10% となった場合, 「とても高い値です。」と音声化する. この際, 音声の高さを通常の 1.5 倍に設定する. また, 値が下位 10% は音声の高さを通常の 0.75 倍にする. そしてライブデータの比較結果が最大の値であった時, 「最大の値です。」と音声化するが, この時の音声の高さは 2 倍にする. また, 最小の値の時は音声の高さを 0.5 倍とする.

● 声量調整部

述語部分は通常よりも音声の大きさを 1.1 倍に設定した.

● 速度調整部

「文章の容器」部分より, その中に入るその場での情報部分は音声の速さを約 1.2 倍遅くして音声化した.

3.2 実装

情報抽出部のプログラミング言語として Java を用い, SOXFire より情報を取得した. その他のサーバ側処理は python3.6 を使用した. また, 実況 lizer は Web アプリケーションとして実装し, Web フレームワークとして django を使用. クライアント側のブラウザ操作として JavaScript を使用した. Ajax のフレームワークとして vue.js を使い, 受信した音声の再生を行なった. また, 情報の送受信には axios を使用した. また, データベースには MySQL を使用した. データベースには後に参照されるために 1 行ずつデータ挿入するテーブルと, 検索速度を考慮して一つのトランスデューサに 1 行データを追加したテーブルを用意した. また, トランスデューサ毎に一つの行を持ったテーブルは, データが流れてくる度にその中身を更新する方法をとった. この行内には今までの同一トランスデューサの値の分布が入っている. 今までの最大値と最小値を格納し, その間を 10 分割した. そしてその値の間にいくつデータ

があるかを記録することにした. 更に, ノード名とトランスデューサ名に複合インデックスを貼ることによって, より早い検索速度が出るように工夫した.

また, 文章化したものを音声化するには VoiceText Web API[8] を用いた. しかし, これを毎度文章を作るたびに使うと, 保存する音声が多くなりすぎてしまう. そこで, 生実況における「(何秒, または何分)」の部分は 1 秒から 59 秒と 1 分から 5 分が収録された音声をあらかじめ用意しておき, その URL をクライアントに渡す処理を行った. これにより, 保存する音声を最小限に留めた. これは他の部分, 「(ノード名) の (トランスデューサ名)」や「(とても (高い・低い) 値です」, 「今までで (最高・最低) の値です」の部分にも用いた. すなわち同じ文章中であっても別の音声ファイルを再生しているが, 連続で再生しているのでその切れ目はほとんど意識されない. VoiceText Web API の仕様では文章中で音の要素を変えることはできないが, 本手法により柔軟に音の要素を変化させている.

4. 評価

4.1 実験概要

実況 lizer が情報の可視化に対して認知負荷を抑えつつ情報提供ができていのかを評価する. 実験では 2 種類の作業 (キーボードタイピングと百マス計算) を用意し, 別々の被験者グループ (タイピング: 20 名 (10-20 代の大学 1 年生), 百マス計算: 11 名 (10-20 代の大学生)) に作業を行ってもらいながら可視化および実況 lizer を用いた情報伝達実験を行った. まず, 両方の被験者グループとも, タイピングおよび百マス計算の練習を行ってもらい, その後可聴化, 実況 lizer 音声工夫なし (声量・声高・速度変更なし), 実況 lizer 音声工夫あり (声量・声高・速度変更あり) の順番で実験を行った.

(1) 可視化: 文章が画面上に表示される. その文章は, 音声 flowed 時と同じ時間だけ表示される. 被験者は作業を行ないながら, 後述する質問票に答えていく. 新しい情報が表示される際に, 「ポーン」という音によってユーザに通知する. 10 秒ごとに提供される情報は更新され, 合計で 6 つの情報が提供される. すなわち, 1 分の作業時間となる.

- (2) 実況 lizer(音声工夫なし): 3.1.3 における音声生成部の音の高さや速さ, 大きさを変えるアルゴリズムを使用せず, 一定の音の高さ, 速さ, 大きさを可聴化を行う. 被験者は作業を行ないながら, 後述する質問票に答えていく. 合計 6 つの情報が音声で伝えられ, 計 80 秒の作業時間となった.
- (3) 実況 lizer(音声工夫あり): 3.1.3 における音声生成部の音の高さや速さ, 大きさを変えるアルゴリズムを使用した音声流される. 被験者は作業を行ないながら, 後述する質問票に答えていく. 合計 6 つの情報が音声で伝えられ, 計 86 秒の作業時間となった.

被験者は, 情報の内容について 4 つの選択肢が書かれた用紙から正答にペンで印をつけてもらい, その正答率を評価する. 情報が現れる順序通りに, 問 1 問 6 までを用意した. 問 1, 問 4 は数字の部分が変わっている, つまり生実況に関する情報が問われる選択肢が用意されている. 例えば, 下記のような選択肢が用意されている.

- ア. 高田市の雨量の値は 5.0 である.
- イ. 高田市の雨量の値は 6.0 である.
- ウ. 高田市の雨量の値は 7.0 である.
- エ. 高田市の雨量の値は 8.0 である.

この中から, 正答に印をつけてもらう. 問 2, 問 5 は過去の同一トランスデューサーと比較した状況分析に関する情報を問うている. そして問 3, 問 6 は現在の同一種類のトランスデューサーと比較した状況分析に関する情報を問うている. これらから, 音声の高さを変えない生実況部分と, 音声の高さを変えて表現した状況分析部分のどちらがより可聴化した際に相手に正確に情報を伝えられるかということの評価することとした. また, 過去情報との比較と, 現在情報との比較でどちらかがより正確に情報を伝えられているかということについても評価することとした. 最後に, 情報提示が終わるまでに達成した作業の 1 分あたりの平均を測り, ユーザの認知負荷への度合いとして評価を行う.

4.2 実験結果と考察

実験結果を, 表 1 に示す. 実験結果より, タイピングの場合は可視化を行った際の方が作業負荷が低い一方で正答率(情報の伝達性)は実況 lizer の方が高く, 百マス計算の場合は可視化より実況 lizer の方が作業負荷および正答率(情報の伝達率)も高いという結果となった. 作業内容によって結果に差が出た理由として, まず作業自体の難しさが考えられる. タイピングの実験は, タイピングに不慣れた非情報系の学生を対象に行った. 被験者にとってタイピングは, 何かを行ないながらできる軽作業ではなく, 神経を集中させなければならないものであった. つまり, 音を聴きながらタイピングを行うことが厳しかった. よって, 目当ての情報が流れてくるまでタイピングをしながら待つということができなかったため, タイピング速度は遅くなっ

たとえられる. 一方可視化の場合は 10 秒間という時間のなかで好きなタイミングで情報を閲覧することができたため, 作業効率が高かったと考えられる. それに対し, 百マス計算は音を聴きながら作業することが可能な程度の作業であったと考えられる. 百マス計算の際の認知負荷については, 可聴化は可視化よりも大幅に低い結果となった. 特に実験が進んでいくにつれて, 可聴化の方が情報を受け取る時の様子がスムーズである印象を受けた. 最初は情報を受け取るまで耳をすませ, 計算問題の手が動かない被験者もあった. しかし時間が経つにつれて構造化された情報に気付き始め, 計算問題を行いながら情報を取り込んでいる様子が観察された. 音声アルゴリズムの有無によって回答数に少々の違いはあったが, これは音声アルゴリズムが無いものから先に実験をおこなった影響もあると考えられ, 誤差の範囲であると考えられる.

一方, 認知した情報の正確さには可視化・可聴化ともに大きな差がないことがわかった. 本実験では, 作業途中で質問に答えてもらう形式をとったため, ユーザは正解の情報を求めつつ情報提示を受けた. したがって情報が流れてきた瞬間, 情報提示される情報から正解を探し出すまで情報に集中できていたと考えられる. 次に, 可視化と可聴化における生実況部分と状況分析部分とを比較する. 可視化において二つに殆ど差はないが, 可聴化では状況分析部分の正答率の方が高かった. これは, すでに与えられている問題文に対し印をつける方式が, いつ答えが流れるかについてのヒントを教えていたことにつながっていたためであると考えられる. なお, 音声工夫の有無によって, 両者の正答率が大きく変わることはなかった. つまり, 正しい情報を伝えられるかという点においては, 実況 lizer の音声工夫の効果が今回の実験では明らかとならなかった.

5. 関連研究

可聴化研究においては, より情報を伝わりやすくする工夫がいくつかある. 例えば DNA の配列を分析する際の補助として音の並びを使うものがある. [10] これはコドンの始まりと終わりをそのまま音にして表現することで, より DNA 配列の周期性がわかりやすくなり, DNA 分析に役立つという研究である. このように, 可視化よりも可聴化の方が直感的な情報提示ができる場合が存在する. 他には, 自然の, 或いは人工の日常にある音を使い, 言葉よりも直感的に人に情報を与えるよう設計されたものとして, 「聴覚のアイコン」(auditory icon)がある. この中に, 雨音の大きさを降水確率が表現されるというものがある. [11] これは降水確率が高いものほど大きな雨音を出力し, 情報を伝えるというものである. 降水確率が徐々に高くなる時は, 徐々に出力する雨音の大きさも大きくしていく. このように, 同じ可聴化という手法をとるとしても, より直感的な情報提示を可能にしている研究もある. しかし, いず

表 1 実験結果

キーボードタイピング		百マス計算	
1分あたりのタイピング数		1分あたり百マス計算量	
可視化	65.3	可視化	24.2
実況lizer(音声工夫なし)	48	実況lizer(音声工夫なし)	49.2
実況lizer(音声工夫あり)	57.8	実況lizer(音声工夫あり)	55
質問表正答率		質問表正答率	
可視化	84.3%	可視化	88.6%
実況lizer(音声工夫なし)	88.2%	実況lizer(音声工夫なし)	90.9%
実況lizer(音声工夫あり)	88.2%	実況lizer(音声工夫あり)	90.9%

れも全体として少ない情報量しか与えられていない。一定の時間で多くの情報を伝えようとする事は、現状の「聴覚のアイコン」を発展させる手法にとっては難しい。

音声化の研究においては、より人間らしく言葉を喋ることに注目した研究は多数存在する。例えば、音声化する際の声の時間構造・強さ・速さから感情を持たせる研究がある [12]。今回の実装に使用した VoiceText Web API も同じく、感情を持たせる機能がある。また、ニューラルネットワークを使い、ナチュラルな発音で英語と北京語とを喋れるようにした研究が存在する [13]。しかし、いずれもより人間と会話ができるように研究されたものであって、情報を伝えることに特化していない。スポーツ実況は必ずしも普段のイントネーションと同じイントネーションで行われるわけではない。視聴者へ情報を正確に、感情を込めて伝える方法論が実況には込められており、その実況のメタファから今回の実況 lizer は実装されている。

6. まとめ

本研究では、都市ライブセンサデータをユーザに伝達するための、可聴化手法実況 lizer の提案を行なった。都市データの可視化だけでなく、効果的な可聴化を実現することで、マルチモーダルな行政職員の都市状況理解を促すことが可能であると考え。本稿では、スポーツ実況のメタファに着目し、その分析と都市ライブデータへの適用を行なった。また、都市ライブデータ流通基盤である SOXFire を活用し、実況 lizer のプロトタイプ実装を行なった。評価の結果、現在の実況 lizer はユーザの置かれる状況、特に軽作業時においては可視化手法より認知負荷を抑えつつ、都市状況の理解を促すことができる可能性が示唆された。今後の課題として、実況テンプレートの拡充、より幅広い表現の実現等を行い、よりユーザ自然に受け入れられるインタフェースの構築を目指す。また、実際の行政職員に利用して頂き、現場での活用における問題点などの発見と解決を目指す。

謝辞

本研究の一部は、国立研究開発法人情報通信研究機構、および国立研究開発法人理化学研究所によって支援された。

参考文献

- [1] ちば市民協働レポート「ちばレポ」
<https://chibarepo.secure.force.com/> (参照 2017-10-09).
- [2] 中澤 仁, 米澤 拓郎, 徳田 英幸, "Sensorizer: ウェブ埋没データからサイバーフィジカルデータストリームを生成するアーキテクチャ", 電子情報通信学会, 信学技報, vol. 114, no. 418, ASN2014-155, pp. 249-254,
- [3] CityDashboard: London
<http://citydashboard.org/london/> (参照 2017-10-09).
- [4] Takuro Yonezawa, Tomotaka Ito, Jin Nakazawa, Hideyuki Tokuda, "SOXFire: A Universal Sensor Network System for Sharing Social Big Sensor Data in Smart Cities", SmartCities' 16 Proceedings of the 2nd International Workshop on Smart Cities, People, Technology and Data, Article No.2 December, 2016,
- [5] Kloeckl, Kristian, Oliver Senn, and Carlo Ratti. , "Enabling the real-time city: LIVE Singapore!", Journal of Urban Technology 19.2 (2012): 89-112.
- [6] 深澤 弘樹, スポーツ実況研究の視座 — 「物語」の視点を中心に—
<http://repo.komazawa-u.ac.jp/opac/repository/all/32747/rsk044-04-fukasawahiroki.pdf> (参照 2017-10-09).
- [7] 山本 浩, "ワールドカップ実況放送の現場から", Journal of mass communication studies (62), 58-81, 2003-01-31.
- [8] VoiceText Web API <https://cloud.voicetext.jp/webapi> (参照 2017-10-09).
- [9] <https://manabi.benesse.ne.jp/gakushu/typing/nihongonyuryoku.html> (参照 2017-10-09).
- [10] Mark D. Temple, "An auditory display tool for DNA sequence analysis", BMC Bioinformatics, 2017
- [11] Stephen Roddy and Dermot Furlong, "Rethinking the Transmission Medium in Live Computer Music Performance", Irish Sound Science and Technology Convocation, Dun Laoghaire Institute of Art, Design and Technology
- [12] Geza Nemeth, Gabor Olaszy, Tamas Gabor Csapo, "SPEMOTICONS:text-to-speech based emotional auditory cues", International Conference on Auditory Display 2011
- [13] WAVENET: A GENERATIVE MODEL FOR RAW AUDIO(arXiv:1609.03499v1[cs.SD] 12 Sep 2016) (Google DeepMind, London, UK) <https://deepmind.com/blog/wavenet-generative-model-raw-audio/> (参照 2017-10-09).