

複数の映像提示方法を備えた オフィス状況伝達システム

比嘉恭太[†], 石川真澄[†], 野村俊之[†], 仙田裕三[†]

複数の映像提示方法を備えたオフィス状況伝達システムを試作した。本システムは、オフィス全体の状況を把握するためのマップモード、オフィスの一部の状況を把握するための実写モード、異なるオフィスに在籍するユーザの状況を把握するための仮想オフィスモードで構成される。これらのモードは目的に応じてインタラクティブに切り替えられる。マップモードと実写モードでは、オフィス状況を直感的に伝えるために発生イベントを擬音語・擬態語で表示する。仮想オフィスモードでは、ユーザの作業集中度を伝えるために単位時間あたりの打鍵数とマイク収録音量をグラフ表示する。簡易な実証実験によりこれらの効果を確認した。

A System for Transmitting Office Situation with Multiple Video Presentation Methods

Kyota Higa[†], Masumi Ishikawa[†], Toshiyuki Nomura[†] and
Yuzo Senda[†]

This paper proposes a system for transmitting office situation with multiple video presentation methods. The proposed system consists of three modes: "map mode", which is for grasping a situation of entire office, "live-action mode", which is for understanding a situation of a part of office, and "virtual office mode", which is for browsing situation of co-workers who are at different locations. These modes are switched over freely depending on user's purpose. In the "map mode" and the "live-action mode", this system uses onomatopoeia as intuitive expression of events occurring at the office. In the "virtual office mode", this system plots the number of keystrokes and the sound volume every second for each co-worker to present his/her preoccupation with work. A simple demonstration result shows that this system is effective to understand an office situation immediately.

1. はじめに

ワークライフバランスの改善や省エネルギー化といった社会的課題を解決する手段としてテレワークが注目されている。テレワークとは情報通信技術を活用した場所や時間にとらわれない柔軟な働き方とされている。近年の情報通信技術の向上により、テレワークを支える多様な機能が実現されつつある。例えば、ネットワークのブロードバンド化やシンクライアントの普及などにより自宅や外出先から社内ネットワークにセキュアにアクセスする環境が整備され、オフィスに出向かなくとも業務に必要な電子情報を共有できるようになっている。また、テレビ会議システムやメッセージなどのコミュニケーションツールも用意され、必要に応じて同僚とコミュニケーションをとることも可能になっている。

一方、オフィスに流通している情報は必ずしも電子化されている訳ではない。フェイス・トゥ・フェイスで行われている会話の内容はもちろん身振りや声の抑揚などの非言語情報は、実際にオフィスにいないとわからない情報である。このような情報には個人の状況やオフィスの雰囲気や推測する手がかりが含まれていることが多く、業務を円滑に遂行するには欠かせない情報となっている。しかしながら、現在のテレワークシステムには、電子化されていない情報の伝達についてあまり考慮されていない。そのため、周囲の状況が分からない、状況からの気付きの機会が失われる、疎外感を感じるという問題が報告されている¹⁾²⁾。また、相手の状況が分からないために、用事があるのにコミュニケーションをとることを遠慮するケースも発生している。

このような問題に対して、全員がオフィスにいるように、もしくはそれ以上に効率よく、コミュニケーションができる超臨場感テレワークシステムが検討されている³⁾。このような臨場感を越えるテレワーク環境を実現するために、我々はこれまでにオフィス環境から取得した大量の視聴覚情報を忠実に再現するだけでなく、ユーザの意図に基づいて視聴覚情報を制御する機能を開発してきた⁴⁾⁵⁾。

しかしながら、通常の業務を遂行しつつ、常に遠隔オフィスの映像を視聴するのは困難である。そのため、オフィスの雰囲気や個人の状況を知りたいときには、その前後、数十秒から数分のオフィス映像を視聴する必要がある。このため、オフィスの雰囲気や個人の状況の把握に時間を要してしまい、円滑なコミュニケーションの妨げになるという問題がある。

そこで本稿では、複数の映像提示方法を用いて、ユーザがオフィスの雰囲気や個人の状況を知りたいときに、一目見てその状況を把握できるシステムを試作する。

以下本稿では、2章で状況伝達システムの要件と実現方法について、3章でオノマトペによる状況提示について、4章で状況伝達システムの構成について、5章でデモ展

[†] NEC 情報・メディアプロセッシング研究所
Information and Media Processing Laboratories, NEC Corporation

示と考察について述べる。

2. 状況伝達システムの要件と実現方法

2.1 システム要件

本システムは、遠隔オフィス間のコミュニケーション活性化のためにオフィスの雰囲気やコミュニケーションをとりたい相手の状況の把握を目的とする。オフィスの雰囲気として、例えば、オフィスのこの周辺に人が集まって議論している、人物 A と人物 B が熱心に会話しているなど、オフィス内の大域的な状況や局所的な状況を確認できることが求められる。また、コミュニケーションをとりたい相手の状況としては、コミュニケーションの可否を判断できるように、相手がどの程度作業に集中しているかを把握できることが求められる。さらに、業務を遂行しながらこれらの情報を確認可能な提示方法が求められる。例えば、オフィスでの会話や音響的なイベントを確認するためにオフィス音をスピーカ再生した場合、常に音が聞こえるため自身および周囲の業務の妨げになる。また、会話内容が全て知られることに抵抗感を持つ人もいる。したがって、音響再生が不要な確認方法が求められる。また、映像に関しても、変化の少ないオフィス映像の状況を一目瞭然に認知できる提示方法が求められる。

以上を踏まえて、オフィスの状況伝達システムに求められる要件は以下 3 つとなる。

- 要件 1. オフィスの大域的な状況や局所的な状況を把握できること
- 要件 2. コミュニケーションをとりたい相手の集中度を把握できること
- 要件 3. 業務の遂行を妨げずに瞬時に状況を把握できること

2.2 実現方法

2.1 に記載した要件 1 と要件 2 を満たすために、マップモード、実写モード、仮想オフィスモードの 3 種類の映像提示方法を設計する。マップモードと実写モードは要件 1 に、仮想オフィスモードは要件 2 に対応している。以下に各モードの概要を記載する。

・ マップモード

オフィス全体を示す俯瞰図上に、オフィス全体で観測された音響情報と映像情報を提示する。これにより、ユーザはオフィス全体の状況を把握できる。

・ 実写モード

ネットワークカメラで撮影したオフィスの実写映像上に、撮影範囲で観測された音響情報と映像情報を提示する。これにより、ユーザはオフィスの一部の状況を詳細に把握できる。

・ 仮想オフィスモード

複数の異なるオフィスに在籍する人物の観測情報を提示する。これにより、ユーザはコミュニケーションをとりたい相手の作業状況を把握できる。

2.1 に記載した要件 3 を満たすために、マップモードと実写モードでは、オフィスの音響情報と映像情報を人が直感的に理解しやすいオノマトペ（擬音語や擬態語）⁶⁾に変換してテキスト表示する。擬音語表現により、スピーカ再生で業務が妨げられることなく、かつ、会話内容を秘匿化した状態で音響情報を把握できる。また、擬態語表現により、時間経過に伴うオフィスの視覚的な変化を瞬時に把握できる。一方、仮想オフィスモードでは、人物の作業状況を知る上で有用なキーボード打鍵数・マウスクリック数とマイク収録音量の推移をグラフ表示する。これにより、作業状況の変化を視認できるため、作業集中度が把握できる。

3. オノマトペによる状況提示

3.1 音響情報の擬音語変換

擬音語変換のブロック図を図 1 に示す。本システムは、音響信号の音量と種別を分析し、種別ごとに用意した擬音語データベースの中から、音量に応じて擬音語を選択する。音量には短区間の振幅エネルギーをリーク積分した値を用いる。音響信号の分類には、ITU-T G.720.1 (Generic Sound Activity Detector)⁷⁾を用いる。これにより、音響信号は、無音・雑音・音楽・音声の 4 種類に分類される。擬音語の選択では、音響信号の種別が「無音」の場合は「シーン」を選択する。音響信号の種別が「雑音」の場合は音量を 2 値化し、値が 1 ならば「ザワザワ」を選択し、値が 0 ならば何も選択しない。音響信号の種別が「音楽」または「音声」の場合は音量を 4 値化し、値が大きい方から順に「ギャガヤ」「ペラペラ」「ボンボン」「ヒソヒソ」を選択する。なお、2 値化または 4 値化した値が大きいほど音量が大きいことを示している。

音響情報のオノマトペ提示例を図 2 に示す。本例では、入力された音響信号は音声として分類され、音量がやや大きいため擬音語「ペラペラ」が選択されている。なお、擬音語を画像中に重畳する位置は、マイクの画像内の位置が既知であるとしてマイクの画像内位置の上方とする。

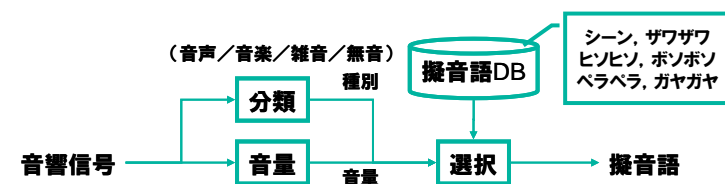


図 1 擬音語変換の処理ブロック図



図 2 音響情報のオノマトペ提示例

3.2 映像情報の擬態語変換

擬態語変換の処理ブロックを図 3 に示す。本システムは、映像信号内の動体を検知し、動体の移動量・移動方向・移動範囲を分析し、事前に用意した擬態語データベースの中から適切な擬態語を選択する。

本システムは以下の手順で動体を検知する。

1. 特徴点の抽出と追跡
入力画像から特徴点を抽出・追跡し、2つの入力画像間の動きベクトルを検出する。なお、特徴点抽出にはハリスのコーナー検出法を、特徴点追跡には Kanade Lucas Tomasi 法⁸⁾を用いる。
2. 動点検出
動きベクトルが規定値以上の特徴点を動点とする。
3. 動点のグループ化
近接する動点をグループ化する。規定数以上の動点が属するグループについて、グループ内の全ての動点を内包する矩形領域を動体として検出する。

擬態語の選択では、動体の移動方向が一定の場合は動体の移動量を3値化し、値が2ならば「スタスタ」を、値が1ならば「テクテク」を選択し、値が0の場合は何も選択しない。なお、3値化した値が大きいほど動体の移動量が大きいことを示している。動体の移動方向が一定でない場合は動体の移動範囲を3値化し、値が0の場合は「ゴソゴソ」を、値が1の場合は「テキパキ」を選択し、値が2の場合は何も選択しない。なお、3値化した値が大きいほど動体の移動範囲が広いことを示している。

本方式による動体検知の一例を図 4 に示す。本方式では、動体の移動量と移動方向は、動体領域内の動きベクトルの平均ベクトルの大きさと方向とする。移動範囲は、動体領域の中心点の軌跡の外接矩形とする。

映像情報のオノマトペ提示例を図 5 に示す。本例では、右上から左下に移動する人物に対して、移動方向が一定であり、移動量が大きいため擬態語「スタスタ」が選択されている。なお、擬態語を画像中に重畳する位置は、どの動体の情報かを明示するために動体領域の上方とする。



図 3 擬態語変換の処理ブロック

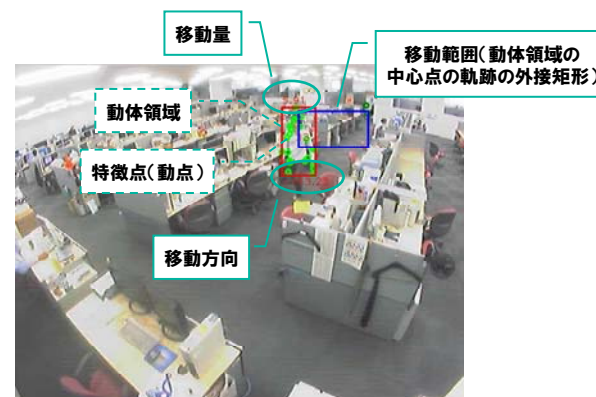


図 4 動体検知の一例



図 5 映像情報のオノマトペ提示例

4. 状況伝達システム

4.1 システム構成

本システムの構成を図 6 に、本システムを構成する主要機器の仕様を表 1 に示す。本システムでは、マイクごとにネットワークオーディオエンコーダを設置しており、ネットワークオーディオエンコーダとネットワークオーディオデコーダは一対一で対応している。また、クライアント端末にはマイクが内蔵された USB Web カメラを設置している。

音響処理サーバは、オフィスに設置されたマイクで収録した音響信号を 10 ミリ秒単位で解析し、1 秒分の解析結果に基づいて擬音語を選択し、選択した擬音語を映像処理サーバに送信する。

クライアント端末は、キーボード打鍵数やマウスクリック数を 1 秒単位で、USB Web カメラで撮影したユーザ画像を 5 秒単位で映像処理サーバに送信する。

映像処理サーバは、オフィスに設置したネットワークカメラで撮影した映像信号 (3fps) を解析し、5 秒分の解析結果に基づいて擬態語を選択し、選択した擬態語を内部メモリに蓄積する。また、映像処理サーバは、音響処理サーバ、映像処理サーバ、クライアント端末の解析結果を用いて、各モードで提示する映像を生成し、クライアント端末の要求に応じて生成した映像を送信する。

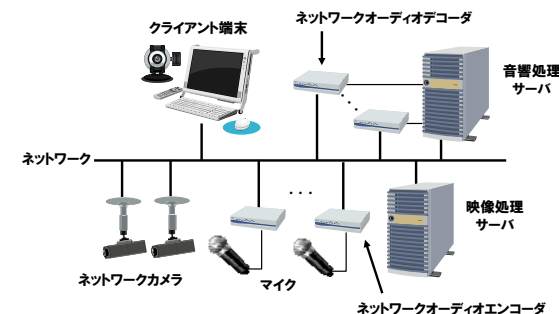


図 6 オフィス状況伝達システムの構成

表 1 状況伝達システムを構成する主要機器の仕様

音響処理サーバ, 映像処理サーバ	HP xw6600 Workstation, Intel® Xeon® CPU E5450 @ 3.00 GHz, 3.25 GB RAM, Windows XP Professional Service Pack 3
マイク	Sony ECM-C10
ネットワークカメラ	Axis 2100
ネットワークオーディオエンコーダ	Barix Instreamer 100
ネットワークオーディオデコーダ	Barix Exstreamer 100

4.2 画面構成

本システムの画面構成を図 7 に示す。①映像表示領域には、各モードの映像 (640x480, 1fps) が表示される。②Tweet 記述領域には、ユーザが自由に文字列を記述することができ、その文字列は③タイムライン表示領域に時系列で表示される。③に表示された文字列は全てのユーザが閲覧できる。④オプション選択領域には、映像表示モードの選択メニューなどがあり、ユーザは目的に応じてインタラクティブに映像表示モードを切り替えられる。

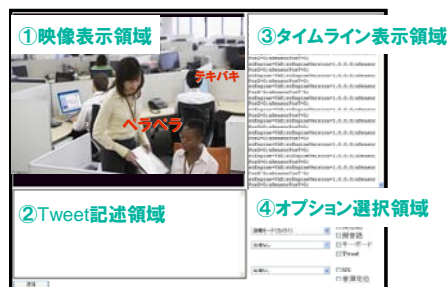


図 7 状況伝達システムの画面構成



図 8 マップモードの提示映像



図 9 実写モードの提示映像の一例

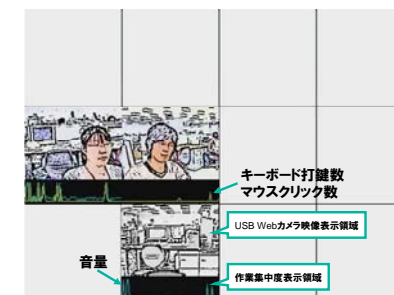


図 10 仮想オフィスモードの提示映像

4.3 マップモード

マップモードの提示映像を図 8 に示す。図 8 の矩形は座席を、黒丸はマイクを、上部と下部の 4 つの図形はネットワークカメラを示している。マップモードでは、オフィスに設置したマイクとネットワークカメラで取得した音響情報と映像情報を擬音語・擬態語に変換し、オフィス全体を示す俯瞰図上に重畳表示する。擬音語はマイク位置の上方に、擬態語は動体領域の位置に重畳する。本システムは、事前に算出したネットワークカメラ映像からマップ画像へのホモグラフィ行列を用いて、マップ画像における動体領域の位置を算出する。

4.4 実写モード

実写モードの提示映像の一例を図 9 に示す。実写モードでは、ネットワークカメラの撮影範囲で取得した音響情報と映像情報を擬音語・擬態語に変換し、オフィスの実写映像上に重畳表示する。擬音語はマイクの上部に、擬態語は動体領域の上部に重畳する。図 9 の提示映像は、図 8 下部の右から 2 番目のネットワークカメラ映像に擬音語と擬態語を重畳したものである。なお、本システムでは、ユーザは注目したい箇所に応じてネットワークカメラを自由に切り替えることができる。

4.5 仮想オフィスモード

仮想オフィスモードの提示映像を図 10 に示す。図 10 は、3 名のユーザが本システムを使用している例である。仮想オフィスモードでは、クライアント端末に設置した USB Web カメラでユーザの上半身を撮影し、そのユーザ画像を仮想オフィス空間に配置する。なお、ユーザ画像の配置位置は事前に指定している。本システムは、プライバシーに配慮してユーザ画像を絵画風に変換する。また、ユーザの在籍状況を瞬時に把握できるように、ユーザが在席している場合はカラー画像を、そうでない場合はモノクロ画像を表示する。なお、ユーザ画像から顔が検出された場合を在席とする。ユーザ画像下部の作業集中度表示領域には単位時間あたりのキーボード打鍵数・マウスクリック数と USB Web カメラに内蔵されたマイクで収録した音の音量が折れ線グラフで表示される。グラフは単位時間ごとに左方向にスクロールし、過去 80 秒間分のデータを表示する。

5. デモ展示と考察

2011 年 6 月 16 日に日本科学未来館で開催された超臨場感コミュニケーション産学官フォーラムのシンポジウムにて、本システムのマップモードと実写モードをデモ展示した。デモ展示では、展示会場に設置したネットワークカメラとマイクを用いた実写モードの動態デモと、下記 3 種の効果を示すデモビデオを展示した。

1. イベント発生への気づき

大勢の人がオフィスの一角に集まり、連絡事項が伝えられているシーン。人が徐々に集まることによる擬態語の増加、および連絡事項を伝える声と周囲のざわめきを捉えた擬音語の発生により、ユーザはオフィスで起きたイベントを瞬時に把握できる。

2. コミュニケーションをとりたい相手の状況把握

コミュニケーションをとりたい相手が別の人物と座席に座って会話しているシーン。2人の会話が擬音語として提示されるため、ユーザは遠隔地にいながらコミュニケーションをとりたい相手が、今すぐコミュニケーションをとれる状況であるか把握できる。

3. コミュニケーション要求に対する気づき

1人の人物がオフィスに不在のユーザの座席に訪れて座席の様子を見るシーン。ユーザに対して用事がある訪れた人物の動きが擬態語として提示されるため、ユーザは遠隔地にいながら自身へのコミュニケーション要求に対する気づきが得られる。

デモ展示を通じて、遠隔オフィス間のコミュニケーション活性化に向けた取り組みとしてデモシステムのコンセプトは概ね賛同が得られた。また、オノマトペ利用に関してエンタテイメント系での利用を推奨するコメントが多々得られた。以下に、来場者から得られたコメントの一部を記載する。

- ・ パラ言語（話し手が聞き手に与える言語情報のうち、イントネーション、リズム、ポーズ、声質といった言語の周辺の側面）を組み合わせると面白そう
- ・ エンタテイメントへの応用が有効
- ・ オフィスの状況を的確に伝えるだけでなくプライバシーを保護する点も特色に思える
- ・ （デモビデオを見て）人が集まり会話しているシーンに対して適切なオノマトペが選択されている

6. おわりに

本稿では、複数の映像提示方法を用いて、ユーザが一目見て的確にオフィス状況を把握できる状況伝達システムを試作した。本システムは、オフィス全体の状況を把握するためのマップモード、オフィスの一部の状況を把握するための実写モード、異なるオフィスに在籍するユーザの状況を把握するための仮想オフィスモードで構成される。これらのモードは目的に応じてインタラクティブに切り替えられる。マップモードと実写モードでは、オフィス状況を直感的に伝えるためにオフィスの音響情報と映像情報を擬音語・擬態語で提示する。仮想オフィスモードでは、ユーザの作業集中度を伝えるために単位時間あたりのキーボード打鍵数・マウスクリック数とマイク収録音量をグラフ表示する。超臨場感コミュニケーション産学官フォーラムのシンポジウムでのデモ展示を通じて、遠隔オフィス間のコミュニケーション活性化に向けた取り

組みとしてデモシステムのコンセプトは概ね賛同が得られた。また、オノマトペ利用に関してエンタテイメント系での利用を推奨するコメントが多々得られた。

謝辞 本研究の一部は独立行政法人情報通信研究機構からの委託研究「革新的な三次元映像技術による超臨場感コミュニケーション技術の研究開発」による。

参考文献

- 1) 日本テレワーク協会, “テレワークの推進のための調査研究 報告書”, 総務省請負事業, Mar. 2007.
- 2) 総務省, “平成 20 年度 テレワーク普及促進のための調査研究”, Mar. 2009.
- 3) 超臨場感産学官フォーラム 技術開発部会 五感・認知分科会 テレワークのための超臨場感技術作業班, “テレワークのための超臨場感技術開発の提言”, Mar. 2009.
- 4) 比嘉恭太, 野村俊之, “分析バッファ制御を用いた Block-Wise ICA によるリアルタイム音源分離”, 電子情報通信学会総合大会, A-4-15, Mar. 2010.
- 5) 野村俊之, 比嘉恭太, 仙田裕三, 芹沢昌宏, “超臨場感テレワークシステムにおける能動的メディア処理の実現に向けた音源分離定位品質の基本評価”, 応用音響研究会, Mar. 2010.
- 6) 荻坂直行, “感性のことばを研究するー擬音語・擬態語に読む心のありか”, 新曜社, Jul. 1999.
- 7) Recommendation ITU-T G.720.1: Generic sound activity detector, ITU-T, Jan. 2010.
- 8) Jianbo Shi and Carlo Tomasi, “Good Features to Track,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 593-600, 1994.