

# ユビキタスマーティングからの マルチモーダル知識獲得に関する研究

山本剛 坂根裕 竹林洋一  
静岡大学情報学部

## 1 はじめに

普段の何気ない会話。友人との会話、上司との会話、家族との会話。こういった会話の中には雑学や生活の知識、自分にとって有益な情報、友人と共有することで楽しめる情報など、様々な情報が含まれている。

ウェアラブルコンピュータを初めて身に着けた Steve Mann[1] は、人生アルバムを作ることがその目的のひとつであったが、人と人とのコミュニケーションにおいて、ユビキタスヘッドセット等を利用した情報収集と構造化は、さらなる価値を生み出すことができる。

本研究では、ウェアラブル機器を常時着用し、人の意図的・非意図的・非意図的情報がある程度理解することにより、人と人との日常会話に内在する意味のある部分を抽出し、理解しやすい形としてコンテンツ化する方法を提案する。

## 2 マルチモーダル情報による会話の要点把握

会話は必然的に発生するとは限らない。偶然発生した会話の中にも大切な情報が含まれていることが多く、今までこの情報を電子データとして記録することは不可能であった。

ウェアラブルコンピュータを利用することで、常時事物を記録することが現実的なものとなり、突然発生する会話でも記録できるようになった。しかし、ストレージ容量や、CPU・ネットワーク速度の制約はなくなっても、それだけでは蓄積情報が膨大になるだけで、どこに重要な情報があるのか把握できない。

重要な情報を把握するための技術として、人間の自然動作を考える。人間が誰か他の人と会話をする際には、言語のコミュニケーションと同時に、相手

の身振り・表情・声の調子などといった自然な動作を観察することで会話が成立している。こういった自然な動作をする際には、何かしら自分が納得したり、相手に気持ちを伝えたいところであったりと、長い会話の中でも重要な箇所である可能性が高い。

本研究では、会話中の自然動作をウェアラブル機器に搭載したセンサによって認識し、話者の声の調子・表情・ジェスチャーなども映像・音声として記録する。こういったマルチモーダル情報を統合的に処理することで、重要な箇所を把握するだけに止まらず、理解しやすいコンテンツとして再構築し提示することを目指している。

## 3 システムの概要

### 3.1 システムの流れ

本研究で提案している、ユビキタスマーティングに内在する知識を抽出する過程を説明する。

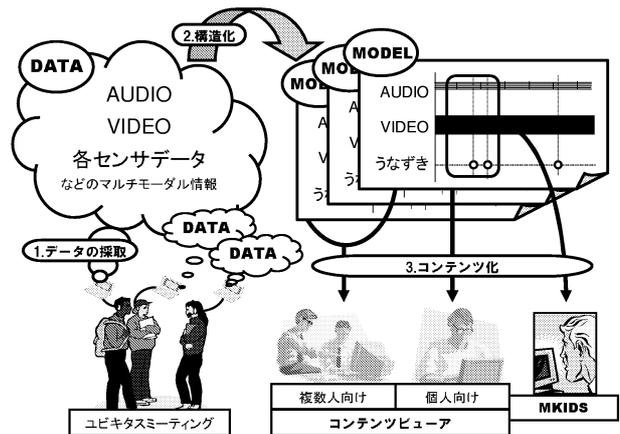


図 1: ユビキタスマーティングからのマルチモーダル知識抽出

Research on the multimodal knowledge acquisition from a ubiquitous meeting.  
Goh Yamamoto, Yutaka Sakane, Youichi Takebayashi  
Faculty of Information, Shizuoka University

会話からデータ採取 カメラ・センサなどを搭載したヘッドセット型ウェアラブル機器を装着し、普段の会話から視点映像・音声・頭の動きをセンシング

したデータなどを採取する(図1中「1.データの採取」参照).

**データの構造化** 音声データの中から特殊な発言を認識したり, センサデータから特定の動作を認識する. そういう抽象化した情報や映像・音声などのデータを, 時系列に並べ, 構造化する. ここで参加者人数分のモデルを作成する(図1中「2.構造化」参照).

**自然動作をした箇所を中心に知識抽出** 構造化されたモデルから, 重要な箇所を **MKIDS(Multimodal Knowledge and Information on Demand Service)[3][4]** を始めとするコンテンツ提示システム [5] 用にオーサリングしてコンテンツ化する. あるいは, モデルから自分に有益な情報をリストアップして自分専用の議事録を作成したり, 会話に参加した他の人のモデルも合わせることで, 会話全体をまとめた議事録を作成する(図1中「3.コンテンツ化」参照).

### 3.2 ウェアラブル機器

本研究では会話データを常時収集するために, ウェアラブル機器を利用する. 今回実装したヘッドセット(図2参照)には以下のものが搭載されている. 図2にその利用画面を示す.

**カメラ** 着用者の視点からの映像を撮る. 話し相手の表情やジェスチャーなどを解析することで効果的なコンテンツを作成する.

**マイク** 着用者の発声を録音する. 発話内容を録音すると共に, 特定の発言をした箇所を把握するためにも使用する.

**センサ** 着用者の頭の動きをセンシングする. センサの種類は 3軸ジャイロセンサ・2軸角速度センサ・地磁気センサの 3種類(図3参照).

ヘッドセットを着用することで, 常に視点映像と



図 2: ヘッドセットを装着した会話風景

自分の音声を記録し, 頭の動作を認識する. また, 音声認識機能を備えたユビキタスヘッドセット [2] を使用することにより, 特定の発言を認識することも可能となる.

## 4 まとめ

カメラとセンサを搭載したヘッドセットを用い, 会話中のうなずきや首振りデータを会話コンテンツに付与することで, 効果的な知識獲得やコンテンツ再構成が行えることが分かった. センサデータの解析を進め, 非意図的なうなずきや行動が詳細に認識できれば, コミュニケーションの理解が深まり, より効果的なコンテンツ生成が可能であるという見通しが得られた.

## 謝辞

本研究の一部は, (株) 東芝の協力による.

## 参考文献

- [1] Steve Mann: Smart Clothing: The Wearable Computer and WearCam, Personal Technologies, vol.1 (1997).
- [2] 金澤博史, 友田一郎, 高島由彰, 竹林洋一: ユビキタス社会に向けた Bluetooth ヘッドセットの開発, 日本音響学会 2002 年春季研究発表会, pp.219{220 (2002).
- [3] 鈴木優, 岐津俊樹, 宮澤隆幸, 浦田耕二, 綱淳子, 竹林洋一: マルチモーダルナレッジをオンデマンドで配信する MKIDS システムの開発, 人工知能学会全国大会, 2D1{03 (2002).
- [4] 中山康子, 真鍋俊彦, 竹林洋一: 知識情報共有システム (Advice/Help on Demand) の開発と実践: 知識ベースとノウハウベースの構築, 情報処理学会論文誌, Vol.39, No.5, pp.1186{1194 (1998).
- [5] Thad Starner: Remembrance Agent: A continuously running automated information retrieval system, Proceedings of The First International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology (PAAM '96), pp.487{495 (1996).

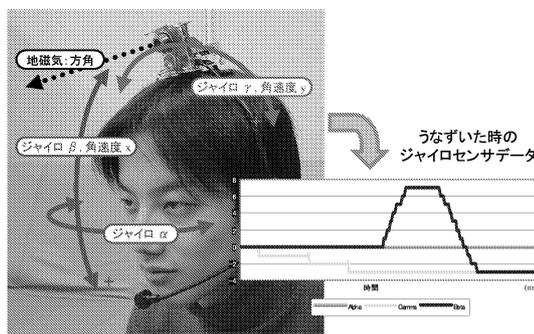


図 3: ヘッドセットのセンサ情報