

「君を幸せにする」と言われても、声に覇気がなかったり目が泳いでいたりすると信じていいの不安になる。ヒトは発言内容（言語情報）に加えて、それを言うときの様子、すなわち声の調子（韻律）、表情、視線、身振り、身体接触などといった複数の異なる器官で得られる感覚（モダリティ modality）を統合して発言や場の理解・推論に役立てる。このように複数（multi-）のモダリティを統合することをマルチモダリティとよぶ。マルチモーダル（multimodal）はその形容詞形である。

よく似たことばに「マルチメディア」がある。マルチメディアとマルチモダリティとは密接に関係する概念であるが、その内容に違いがある。マルチメディアは媒体（media）に注目した概念で、たとえば音楽の流れる絵本のように文字、画像、音声といった異なる複数の情報提供媒体を組み合わせることである。一方、マルチモダリティは知覚に関するものであり、同じ例でいえば視覚では文章と絵、聴覚では音楽と読み聞かせの声を知覚し、それらを統合して作品を体験することに該当する。

「マルチモダリティ」の語の使われ方には分野間でずれが見られる。ここでは、マルチモダリティの概念が登場した工学と人文社会系を例にみていきたい。前者では、ヒトと機械（コンピュータ）とのインタフェースに対して使用される。従来のキーボードによる文字入力に対して、現在のスマートフォンなどではマイクやカメラ、センサーを用いた音声・画像・指の動きなどによる入力が可能なマルチモーダル・インタフェースも搭載されている。

マルチモダリティ Multimodality

かおだ じゅんぺい
金田 純平 民博 機関研究員

君を幸せにする
人間学の
キーワード

る。これはヒトのマルチモーダルな知覚に倣ったものという意味であり、ここでいうモダリティはむしろ情報獲得装置あるいは手段という意味である。

後者では、ヒト同士の会話ややりとり（インタラクション）の研究において、発言内容の記録を超えて、韻律、表情、視線、身振りといった非言語的特徴にも注目した記述・分析をおこなうことをマルチモーダルな記述とよぶ。話すという日常的な行動に潜む、普段ヒトが無意識的におこなっていることを意識化・可視化させる手法である。この場合のモダリティは、知覚可能で何らかの意味や表象を持つと考えられる言語・韻律および各種非言語行動（視線など）の区分を指す。

工学および人文社会系の分野で使用される「マルチモダリティ」に共通することは、複数の装置を用いて異なる特徴を持つデータを得て、それらを統合し利用することである。また、非言語のデータを重視するというポイントが肝要である。言語によるコミュニケーションは抽象的で複雑な内容を処理できるが、そのぶん内容理解のための認知的・時間的コストが大きくエラー（聞き違い）も発生しやすい。一方、感覚知覚や直観的操作（たとえばマウスのクリックやタッチパネルでのスワイプなど）は認知的負担が軽く短時間で済む。マルチモダリティにかかわる研究とは、カメラなどの機器を複数種類用いて、言語情報と対比させながらヒトの直観的な認知や感情、無意識について明らかにしていくことである。