要旨

近年、コンピュータやロボットなどの機械が人間の生活において担う役割は、ますます重要なものになりつつある。これに伴って、より柔軟な人間的ヒューマンインターフェースに対する要求が高まりつつある。

人間同士のコミュニケーションにおいては、非言語的情報の重要性は古くから知られており、その中でも顔表情や音声等に自然に表れる「感情」は非常に重要な役割を果たしている。Mehrabin によると人間のコミュニケーションにおいてメッセージのわずか 7 %が言語で伝達される一方で、55 %は顔の表情によって伝達されるとしている。このように、人間とコンピュータの自然なコミュニケーションを実現するためには、人間に負担をかけない非接触方式により人間の感情を認識する事が重要な課題である。

人間がコミュニケーションにおいて、顔表情や音声の両者から相手の感情を認識しているにもかかわらず、従来の研究では顔画像単独や音声単独からの感情認識が大半である。そこで、本研究では感情認識において画像情報と音声情報を統合することで、認識率の向上を行うと共に、画像情報や音声情報の劣化の激しい悪条件な環境に強い感情認識を行った。

本研究では画像情報として眉、目、口といった顔部位の特徴点の座標を、音声情報として発話音声のピッチの時系列情報を使用し、画像情報単独による認識率、音声情報単独による認識率、画像情報と音声情報の統合によって得られた認識率を調査・比較した。その結果、画像情報と音声情報との結果統合において認識率の向上が見られた。

また、画像情報や音声情報の劣化が激しい悪条件な環境に強い認識を行うため、画像情報を劣化させたデータ、音声情報を劣化させたデータによる認識率を調査・比較した。その結果、一方の情報が劣化した際にでも、他方の情報単独による認識率を確保できた。