

メロディによるロボットとのコミュニケーション

真下奈緒子 中川弘隆 藤井正範 藤波努
北陸先端科学技術大学院大学 知識科学研究科
{n-mashi, hirotaka, m-fujii, fuji}@jaist.ac.jp

1. はじめに

近年、コミュニケーション・ロボットやエンタテイメント・ロボットが製品化され、ロボットと接する機会が増えてきている。また、文書作成やソフトウェア操作を音声で行うことが可能になり、ロボットを音声で操作することも行われている。ここで用いられる音声認識の多くは、音声によって入力された言語パターンを、無数にある言語パターンから特定するものである。使用者によってアクセントや声質に違いがあるため、前もって使用者の声を登録しておく必要がある。しかし、同じ使用者であっても体調や心身状態による声質の変化には対応できないため、いつでも同じように特定するのは困難である。

本研究では、音声装置から出力されるサウンドデータの音階を特定し、9音階からなる連続した3音をメロディとして認識するシステムを製作した。これにより、音質の変化を気にすることなく、人間と自律移動型ロボット GAIA-2 が 504 通りのメロディパターンによってコミュニケーションをとることを可能にした。なお、9音階のほかにもシステム用として3音階を用いている。

2. コミュニケーションとは

人間とロボット、つまり機械とのコミュニケーションを考えるにあたり、何を以ってコミュニケーションが取れているとするのが問題となる。ミード[3]によれば、コミュニケーションとは「記号を媒介とした相互作用過程」であり、記号とは身振り、表情、ことば、文字、映像などのことである。さらに、コミュニケーションは送り手の記号を受け取った「受け手の反応」によってその意味が決定されるものであり、コミュニケーションは原則的に「相互作用」によって成立するものである。

今回、我々は「記号」としてメロディを用いた。人間がロボットにメロディを送信すると、ロボットが行動によって反応する。逆にロボットが人間に信号を送信すると、人間がメロディで答えることで反応する。このような相互作用が行われることでコミュニケーションが取れているものとした。

メロディを用いるメリットには次のようなものがある。

- 音声言語に近いので、使用者にとっても傍観者にとってもコミュニケーションの媒体として用いられていることを認識しやすい
- ロボット側での処理が速い
- 人間とロボットの双方にとってパターン認識が容易である

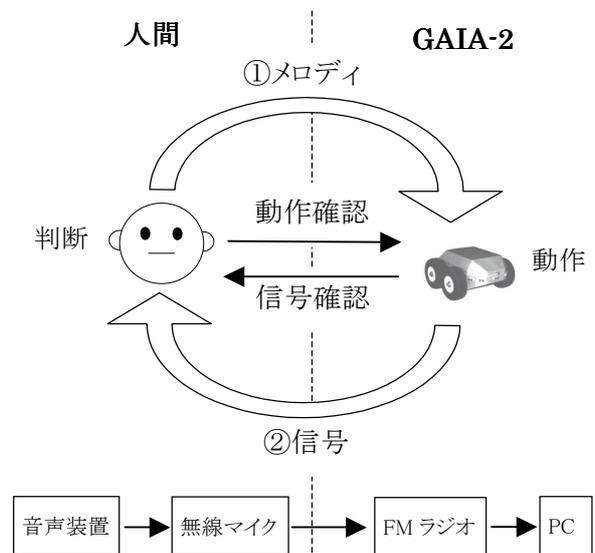


図 1: 動作説明図

3. 人間とロボットとの相互作用過程

コミュニケーションを行う目的は、メロディや行動によって、人間と GAIA-2 が意味を共有できるメロディを見つけることである。

初期状態では、GAIA-2 にはメロディに対応する行動がプログラミングされていない。しかし、システム用の 3 音階は識別できる。以下に動作内容を示す(図 1)。

Communication with the robot using melody

Naoko Mashimo, Hirotaka Nakagawa, Masanori Fujii and Tsutomu Fujinami

Japan Advanced Institute of Science and Technology

ランダムな動作を繰り返す GAIA-2 に、人間がメロディを送信する(①)と、GAIA-2 上のメロディ識別システムが起動してコミュニケーションが始まる。

はじめ、GAIA-2 には記憶されているメロディが無いので、ランダムに動作する。人間はこの動作を確認しておく。動作後、GAIA-2 は人間の期待する動作を行ったかを問う(②)。人間は、期待していたものであれば“Yes”，そうでなければ“No”，もう一度動作して欲しい場合は“Repeat”に対応するサウンドを発信する(①)。“Yes”を受信した場合、GAIA-2 はその時のメロディと動作を記憶しておき、次に同じメロディを受信したときに同じ動作をするようになる。

また、人間の送信したメロディを GAIA-2 が正確に識別できなかった場合、GAIA-2 は人間に、再度メロディを送信するように要求する(②)。人間はこれを受けて再びメロディを送信する(①)。

4. メロディ識別システム

4.1 外部音声入力

外部音声は、音声装置に取り付けた無線マイクにより FM ラジオを介して PC にライン入力される(図 1)。入力された外部音声のアナログデータは、Linux の DSP デバイスを用いて 8000 個/秒サンプリングすることによってデジタルデータに変換される[1]。この結果、音階特定に必要な数値データが得られる。

4.2 音階特定とメロディ識別

音階の特定は、音階によって周期が異なることを利用している。100ms 単位で得られた外部音声のサンプリングデータを解析し、はじめのピーク値となる数値データから、次のピーク値の数値データまでのデータ数をピッチとして求める(図 2)。ピーク値を求める際にノイズの影響を考慮し、求めたピーク値間のデータ数が 7 個以下であった場合、ピーク値として認めない。同様に求めた 100ms 分のピッチデータの平均を算出することで、平均ピッチを得る。あらかじめ行った音階別のデータ解析により、平均ピッチは音階が半音上がると、ほぼ 0.5 ずつ少なくなることが分かっている。外部音声から得られた平均ピッチと、この平均ピッチを比較することによって音階を特定する。

こうして特定した音階を、3 個続けて保持することでメロディとして認識する。メロディ認識中に YES/NO/REPEAT 信号が含まれた場合や、無音、または GAIA-2 が認識できる音域を外れることで、音階が特定できない状態が 1 秒以上続いた場合はメロディとして認識されない。

また、この音階特定方法は、外部音声に倍音成分が含まれる場合には正しく音階を特定することができない。これを解決するためには、対象の音源の音域を制限することや、ローパスフィルタの遮断周波数をできる限り低くするといった工夫が必要となる[2]。そのため今回は、出力するサウンドデータが 12 音階に限定された音声装置を用いている。

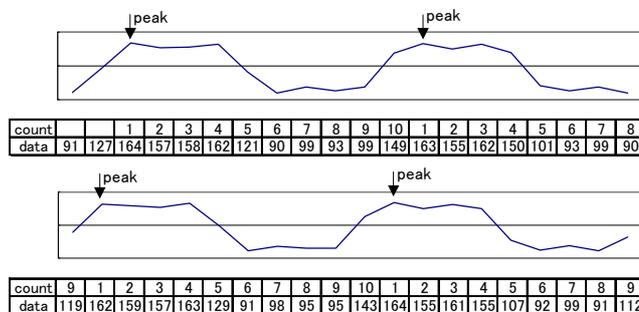


図 2:ピーク値検出

5. おわりに

今回のコミュニケーションでは「記号」をサウンドデータとしたため、12 音階を出力する音声装置を用いたが、このシステムは、入力から音階を特定できるサウンドであれば、どんなものでも用いることができる。今後、様々な楽器や人間の音声等によって GAIA-2 とのコミュニケーションが図れるような仕組みを考える予定である。

また、サウンドを用いたことで、楽しんでコミュニケーションできるという効果も期待できる。

参考文献

[1] Jeff Tranter: LINUX マルチメディアガイド; O'REILLY, 1997
 [2] 片寄晴弘: コンピュータと音楽の世界; 自動採譜(概論); pp.74-pp.88, 共立出版, 1998
 [3] G.H.ミード: 「デューイ=ミード著作集」 精神・自我・社会; 人間の科学社, 1995