

# 音声モンタージュシステムの構築\*

木戸 博 (東北工業大学)

## 1 はじめに

現在の科学捜査において、音声を記録できた場合にのみ、有効な分析をすることができる。しかし、過去の犯罪の解決に、「聴取された音声」が重要な役割を果たした例が少なからずある。つまり、犯人の音声記録されていないときでも、その音声を聞いた人(耳撃者)の記憶に音声は残っており、この記憶(聴取印象)を科学捜査のひとつの手法として活用することが、本研究の背景にある。

本研究では、「音声モンタージュ」を実現するため、要素課題の解明を進め、最終的に実用に値するシステムを完成させることを究極の目的と考えている。

## 2 研究アプローチ

本研究は、被験者が記憶している発話の非言語・パラ言語情報を、記憶内容を歪めずにどのように取り出し、取り出した情報に基づいてどのように合成し、そして、被験者に類似度をどのように評定させるか、という問題に帰着する。

被験者が記憶している音声の非言語・パラ言語情報に関する記憶内容を取り出すには、言語的表現に頼らざるを得ない。そこで「音声モンタージュ」という総合的研究課題を解決するために、以下の6つの要素課題に分けて考える。

要素課題①(音声の表現方法):発話の非言語・パラ言語情報の適切な日常表現語(一般市民が共通概念として持っている特定話者の発話を規定するのに役立つ表現語)リストの構築。要素課題②(表現と音響関連量):日常表現語の音響関連量の解明。要素課題③(音声の記憶):発話の非言語・パラ言語情報の記憶機構の解明。要素課題④(呈示手法):被験者の記憶内容を歪めずに合成音声を繰り返し呈示する戦略の確立。要素課題⑤(類似度評定):合成音声の類似度評定方法の確立。要素課題⑥(柔軟な音声合成):声質や発話様式を柔軟に

制御できる(変換できる)音声合成技術の開発。

以上の要素課題の成果を統合して、システムの構築を目指す。

## 3 各要素課題の研究状況

要素課題④(呈示手法)と要素課題⑤(類似度評定)については、まだ報告できる段階にはないが、他の各要素課題については、現在のところ、以下のような成果を得ている。

### 3-1 要素課題①(音声の表現)

声質表現語について、多数話者から選出した音声資料を用いた聴取評価によって検証し、安定性・信頼性が保証された表現語を抽出した[1,2](表1参照)。また、発話様式など声質以外の表現についても調査を進めている[3]。

表1. 声質表現語対

表現語対	
高い声	— 低い声
男性的な声	— 女性的な声
かすれた声	— 澄んだ声
落ち着いた声	— 落ち着きのない声
迫力のある声	— 弱々しい声
若い感じの声	— 老けた感じの声
太い声	— 細い声
張りのある声	— 張りがない声

表現語(反意語を持たない)  
鼻 声

### 3-2 要素課題②(表現と音響関連量)

非線形解析である決定木を用いて、ほぼ合理的な結果が得られた[4,5](表2参照)。

表2. 抽出された音響関連量

声質表現語	音響関連量
高い	基本周波数, 話速
男性的	基本周波数 SPL, スペクトル傾斜
かすれ	喉頭雑音, ST
落ち着いた	スペクトル傾斜, 基本周波数, SPL, フォルマント構造, 喉頭雑音
迫力	SPL, 基本周波数, スペクトル傾斜, 喉頭雑音, 話速
若い	話速, スペクトル傾斜, 基本周波数, 喉頭雑音
太い	基本周波数, SPL, 話速, スペクトル傾斜
張り	SPL, スペクトル傾斜, 話速, 基本周波数

\* Construction of the voice-montage system, by KIDO, Hiroshi (Tohoku Institute of Technology).

さらに、聴取印象の一つとして欠かせない音声から受ける話者の年齢の印象についても研究を進めている。聴取印象から推測される年齢を「声年齢」と定義し、話者の年齢を音声だけから推定することは、実用上、難しいことを確認した。また、「声年齢」に影響を及ぼす音響的要因についても考察を進めている [6,7] (図1参照)。

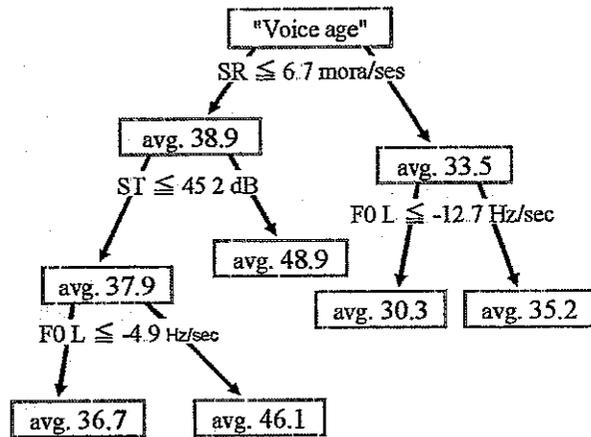


図1. 「声年齢」における決定木分析の結果

### 3-3 要素課題③ (音声の記憶)

再認実験の結果、「特徴的な話者」に関しては3週間程度経過しても記憶の変化が少ないこと、「平均的な話者」ではいくつかの声質表現で記憶が劣化することなどの結果を得た [8,9] (図2参照)。

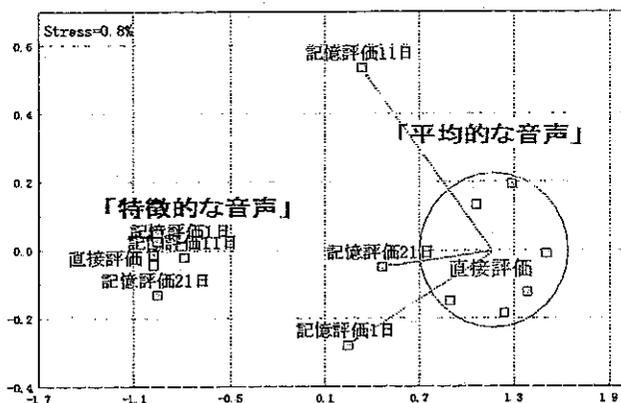


図2. 心理空間における各評価の布置

### 3-4 要素課題⑥ (柔軟な音声合成)

共同研究者の粕谷教授 (宇都宮大) の研究である ARX 分析合成法を用いて、男性はもとより女性・子供を含むいろいろな話者のいろいろな発話スタイルに対して頑健な分析・変換・合成が可能になった [10,11]。

## 4 むすび

「音声モニタージュ」について正面から取り組んだ研究は、国内外ともに、他に見あたらない。それは、「音声モニタージュ」という構想に必要性がないためではなく、音声言語科学・工学・心理学等に関係する総合的研究課題なため、実現するには、克服すべき課題が山積しているからである。

しかし、これは言い換えれば、音声言語科学・工学・心理学の総合研究課題という面で多くの新しい研究分野を提供するものと考えている。本研究を通して、人間の音声記憶機構、発話の声質やスタイルの言語表現法、音声合成における非言語・パラ言語情報の制御手法、個人性のクラスタリング法などの解明あるいは開発につながるものと期待できる。

一方、本研究の成果は、法科学的な面に限らず、将来の高度な音声対話システムを実現する上で必須となる次世代音声合成器において、システム利用者の利用目的・利用環境・利用方法に応じた、柔軟な音声合成システムの要素技術としても貢献できるものと考えられる。

### 参考文献

- [1] 木戸, 粕谷, "通常発話の声質に関連した日常表現語の抽出," 音響学会誌, vol. 55, pp. 405-411, 1999.
- [2] 木戸, 粕谷, "通常発話の声質に関連した日常表現語の抽出 -聴取評価による抽出-, " 音響学会誌, vol. 57, pp. 337-344, 2001.
- [3] 木戸, 粕谷, "音声の聴取印象に関する表現語 -声質・発話様式・好悪-, " 音講論集, pp. 423-424, Mar. 2002.
- [4] 木戸, 箕輪, 粕谷, "声質表現語の音響関連量に関する非線形分析 -決定木による方法-, " 音響学会誌, vol. 58, pp. 586-588, 2002.
- [5] 木戸, 川股, 粕谷, "声質評価システムの構築," 聴研資料 H-2003-67, vol. 33, 2003.
- [6] 木戸, 粕谷, "音声聴取における声年齢についての一考察," 音講論集, pp. 491-492, Sep. 2004.
- [7] 木戸, 粕谷, "声年齢に影響を及ぼす音響的要因," 音講論集, pp. 447-448, Sep. 2005.
- [8] 木戸, 粕谷, "声質の記憶とその時間変化 -音声の特徴の有無による差異-, " 音講論集, pp. 283-284, Sep. 2000.
- [9] 木戸, 粕谷, "声質の記憶とその時間変化 -話者の再認実験-, " 音講論集, pp. 425-426, Mar. 2001.
- [10] Ohtsuka, I. and Kasuya, H., "Aperiodicity Control in ARX-based Speech Analysis-Synthesis Method," Proc. EUROSPEECH, 2001.
- [11] 川股, 粕谷, 木戸, "ARX 音声合成法を用いた声質表現語の音響関連量に関する検討," 音講論集, pp. 289-290, Sep. 2004.