合成音声の韻律的特徴と合成音声品質との関係について *

◎伊藤 嘉治　丁 文　Nick Campbell　樋口 宜男
(ATR 音声翻訳通信研究所)

1　はじめに

ATRで開発した、自然音声波形接続型任意音声合成システムCHATRは、あらかじめ録音された音声データベース中の音素単位の波形を、何らの信号処理も行なわずに接続し、連続音声として出力する [1][2]。そのため、音声データベースを変えることで、話者性や発話様式の特徴を失わずに任意の音声が合成できる。

ところが、異なる音声データベースを使った場合には、たとえ発話内容が同じでも、合成音声の品質にも違いが生じ、好まれる話者、好まれない話者が存在することが経験的に明らかになっている。

そこで、合成音声の持つ音響的特徴量と、合成音声品質の関係を分析し、効率的な合成音声用データベース構築のために必要な音響的性質を解明することを目的とし、本報では、聴取実験による主観評価と、合成音声の音響的特徴量の関係を分析したので、報告する。

2　CHATRの主観評価実験

はじめに、CHATRの合成音声品質を調査するために、聴取実験による主観評価を行なった。

2.1　音声試料

CHATRは音声データベースを変えることにより、様々な言語で、任意の話者の合成音声を合成できる。今回は日本語話者１５名(男性８名、女性７名：アナウンサー、ナレーター、外国人、一般人など)を用い、新聞から選んだ短文１０文を合成した１５０サンプルを音声試料とした。

2.2　実験方法

合成音声の品質を測定するために、音声試料の順番をランダムに変えた５セットを用意し、１５名の被験者(男性８名、女性７名)に各サンプルを１回ずつ提示した。被験者には１セット目には総合評価、２セット目には明瞭性、３セット目には抑揚の自然さ、４セット目にはリズムの自然さ、５セット目には強弱の自然さに着目させ、それぞれ５段階で評価させた。なお、総合評価の際の判断規準は特に示さず被験者の自由な判断に任せた。また、明瞭性の判断は文の聞き取りやすさを評価させた。

2.3　実験結果

主観評価実験の結果を図1に示す。ここでは、被験者による評価値の分布のばらつきを取り除くため、被験者毎に正規化したスコアを用いた。average は各評価値の平均である(図2)。この結果 FMP,FKS, FKT,FYM, MNC の評価が高く、逆に FHS,FKN, FTK, MHN の評価が低かった。

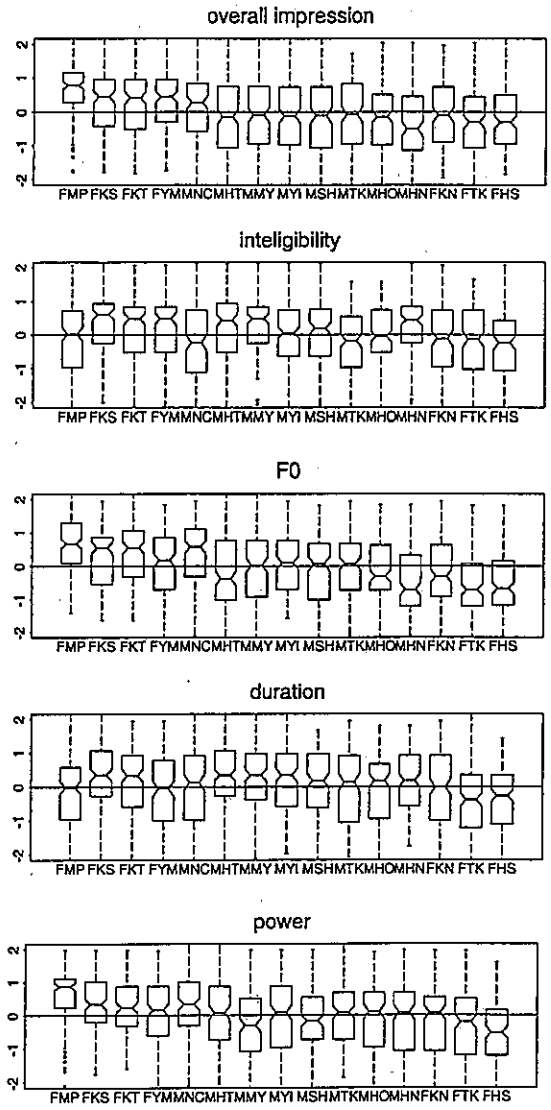

図 1: 評価規準別主観評価値の分布

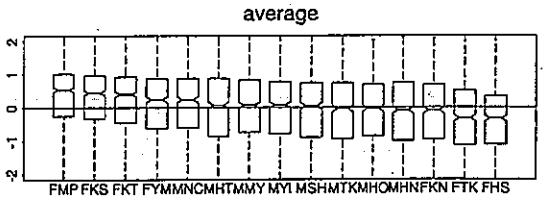

図 2: 平均主観評価値の分布

評価規準間の相関係数を調べると、総合評価は基本周波数と 0.86、振幅と 0.87 で非常に相関が高く、逆に明瞭性、音素継続長との相関はともに 0.12,0.11 と低い (図3)。また、明瞭性は音素継続長との相関が 0.77 と高く (図4)、基本周波数、振幅との相関はとも

*Relationship between Acoustic Characteristics of Synthesized Speech and Quality, by Yoshiharu Itoh, Wen Ding, Nick Campbell and Norio Higuchi(ATR Interpreting Telecommunications Research Labs.)

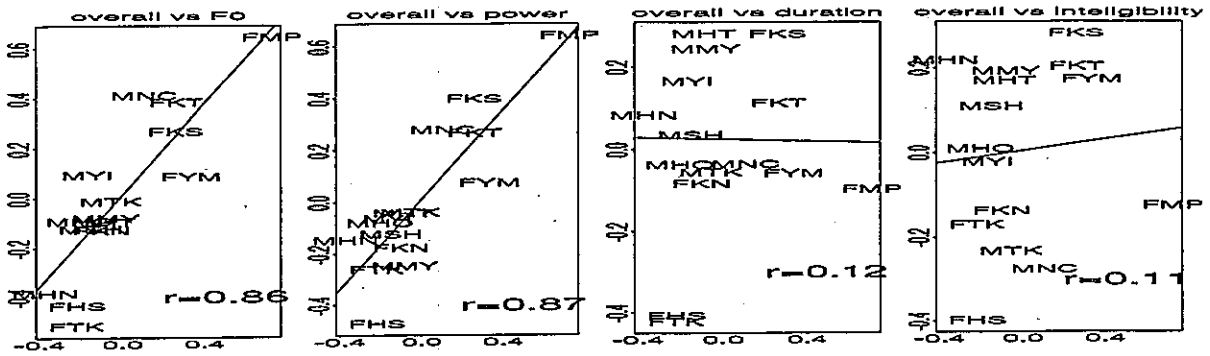

図 3: 総合評価と他の評価規準間の相関

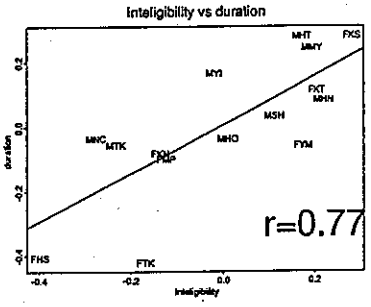

図 4: 明瞭性と音素継続長の相関

に 0.13,0.22 と低い。このことから、総合評価 (合成音声品質に対する好み) では主に基本周波数と振幅が重視され、音素継続長の不自然さには影響されないことが明らかになった。また逆に、合成音声の明瞭性の判断には音素継続長 (リズム) が重視され、基本周波数や振幅の不自然さには影響されないことも同様に明らかになった。さらに総合評価と明瞭性の相関も低く、聞き取りやすい声が、必ずしも高い評価を得るとは限らないことが判った。

なお、明瞭性の平均スコアは他の評価規準に比べ高かった。このことは、原音声の波形をそのまま用いている CHATR の特徴を現していると言える。

3 音響的特徴量との関係

ここでは、総合評価と相関が高い基本周波数および振幅と、主観評価値の関係の分析を行なった。

評価実験に使用した音声試料の基本周波数の分析では、サンプル毎の平均基本周波数の話者別平均および、標準偏差の話者別平均と主観評価値との間で、相関係数がそれぞれ 0.50,-0.71 と明らかな相関が認められた。すなわち、平均基本周波数が高いほど評価が高く (図 5)、基本周波数の平均標準偏差が低いほど評価が高い (図 6)。このことから、CHATR では男声話者よりも女声話者が好まれ、抑揚はある程度抑えられた声の方が好まれると言える。

なお、話者 FHS, FTK はすべての評価項目でスコアが低かった。この 2 話者の声は非常に耳障りな声で、評価した項目以外の要素 (声質や録音状態) が影響していると考えられたため、今回の分析では除外した。

また、振幅については明らかな相関関係は認められなかった。

4 むすび

自然波形接続型任意音声合成システム CHATR の合成音声品質の主観評価実験を行ない、総合評価は基本周波数および振幅の評価と非常に相関が高く、基本周波数および振幅に影響されるとともに、明瞭性は音素継続長の評価との相関が高く、音素継続長に影響を受けていることが明らかになった。また、明瞭性は総合評価に影響しないことも判った。

さらに、合成音声の評価値と平均基本周波数およびその標準偏差との間に相関が見られ、平均基本周波数が高いほど評価が高く、基本周波数の標準偏差が低いほど評価が高いことが判明した。

今後は、合成素片の抽出元となる音声データベースの音響的特徴と合成音声品質の関係や、ターゲットの音素環境と、抽出元の音素環境との関係、スペクトル情報との関係など、本研究で考慮していない要因との関係を分析していく予定である。

また、CHATR は多言語に対応した音声合成システムであるので、英語など他の言語の場合にも同様の関係があるのか検証する必要がある。

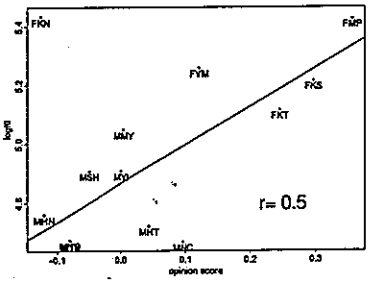

図 5: 平均基本周波数と評価値の関係

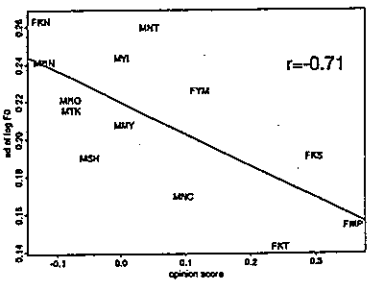

図 6: 基本周波数の標準偏差と評価値の関係

参考文献

[1] W.N.Campbell, A.W.Black, "CHATR: 自然音声波形接続型任意音声合成システム", 信学技報,SP96-7 (1996, 5)

[2] A.W.Black, W.N.Campbell, "Optimising selection of unit from speech databases for concatenative synthesis", Eurospeech'95, pp.581–584 (1995, 9)