

合成音声の高さの弁別閾: 基本周波数の変化による違い*

古山竜司 上田和夫 (九州芸工大)

1 はじめに

短期記憶における音声と非音声の干渉効果を調べる実験で、非音声の高さの再認に対する音声の干渉効果が小さいことを根拠として、音声と非音声の記憶の特殊化について論じられている (Deutsch 1970)。一方、上田、瀬尾 (1998)、上田 (2002) は、音声分析合成法 STRAIGHT (Kawahara *et al.*, 1999) によって、音声の高さを操作し、音声の高さを定常な高さにすることで、調波複合音の高さの再認に対する干渉効果は調波複合音と同様の効果があることを示した。つまり、自然音声の非定常な音を安定した高さに置き換えれば、調波複合音と同じような高さの明確さをもつのではないかと考えた。

そこで彼らは、有声区間の基本周波数が調波複合音と厳密に等しく、高さが定常でかつ音声としての明瞭性も最大限保たれた刺激を作成して実験を行い、音の高さの弁別閾を求めた。その結果、調波複合音と基本周波数を定常にした合成音声との間で弁別閾に有意な差はみられないことから、基本周波数を定常にすることによって、音の高さが明確になり、高さの判断がしやすくなったと考えた。

したがって、非音声の高さの再認に対する音声の干渉効果が小さいのは、音声と非音声の記憶がそれぞれ特殊化されているのではなく、音声の高さが不明確であるためと考えた。しかし、彼らの音

の高さの弁別閾を求める実験は実験参加者の数が少なく、また実験方法も改善すべき点があった。

本研究では、合成音声の基本周波数の変化が音の高さの弁別閾にどのような影響を及ぼすのかを調べるために、音の高さの弁別閾の測定に関して、上田、瀬尾の実験方法を改善し、より多くの実験参加者について実験を行った。

2 心理実験

2.1 刺激

本研究で用いた刺激は、次の三種類である。刺激の合成、あるいは分析再合成は、すべて UNIX ワークステーション (Silicon Graphics, Indy, R4600PC CPU) で行われた。いずれも計算機上では、12 kHz サンプリング、16 bit 直線量子化のデータとして扱われた。

(1) 調波複合音

調波複合音は基本音成分から順に 8 個の成分で構成されたものであり、持続時間は 200 ms、立ち上がり、減衰時間は 10 ms とした。基本周波数 110 Hz を中心として、6 cent ステップで 21 の刺激を作成した。

調波複合音の各成分の振幅 (合成時) は次のようにして決定した。基本音成分の振幅を A_1 とするとき、第 i 倍音の振幅 A_i を、

$$A_i = 2^{-(i-1)} A_1 \quad (1)$$

となるように合成した。

* Pitch differential limens of synthesized speech: The effect of fundamental frequency change. By Ryuji Furuyama and Kazuo Ueda (Kyushu Institute of Design).

(2) 合成音声 I

自然音声 (1名の男性話者による数詞音声, 「0(ゼロ)」から「9(きゅう)」まで) を STRAIGHT version17 による分析を行い, 自然音声から得られた有声区間に基本周波数を 12.5 cent ステップで全体にシフトさせて 21 の刺激を作成した。

(3) 合成音声 II

(2)と同様に分析し, 合成時に有声区間の基本周波数を定常にし, 110 Hz を中心として, 12.5 cent ステップの 21 の刺激を作成した。

自然音声は ATR 人間情報通信研究所の音声データベース (匂坂, 浦谷, 1992) から取り出されて大きさがそろえられ, およその高さが測定されていた。合成音声 I, II の分析合成は, ワークステーション上の MATLAB 環境で行われた。いずれの刺激も 1 kHz, 75 dB の純音と同じ大きさになるように振幅を調整した。

2.2 実験参加者

九州芸術工科大学の学生 10 名が実験に参加した。実験参加者の聴力はオーディオメーター (RION, AA-77) を用いて検査し, 正常であることを確認した。実験参加者に対して, 全実験時間の半分の時間はボランティアで行ってもらい, 残りの半分の時間に対しては謝金を支払った。

2.3 実験手続き

極限法 (二件法) による測定を行った。ワークステーション上にて, 実験参加者ごとにつくられた刺激系列をもとに刺激を生成し, アダプター (STAX, SRM-1/MK-2 pp) を通じて, ヘッドフォン (STAX, SR-303) で実験参加者の両耳に呈示した。系列の開始点は, 端の 3 ステップからランダムに決められた。

実験参加者は, 試行ごとに比較刺激が標準刺激よりも高いか, それとも低いかを判断し, キーボードのテンキーによって反応

した。実験参加者の反応はモニター (Sony, GDM-17E11) に表示された。試行ごとに比較刺激は 1 ステップずつ変化し, 実験参加者の反応が変わったところで次の系列へと進んだ。

調波複合音と合成音声 II については基本周波数 110 Hz の刺激を, 合成音声 I については高さをシフトしていない刺激を標準刺激として用いた。各標準刺激につき, 上昇系列・下降系列を 2 系列ずつ, さらに標準刺激と比較刺激の呈示順序を入れ替えて, 計 8 系列を 1 ブロックとし, 各刺激について 2 ブロックずつの弁別閾の測定を行った。実験は, 九州芸術工科大学大学院芸術工学研究科応用情報伝達講座の防音室で行われた。防音室内の暗騒音を, 普通騒音計 (小野測器, LA210) で測定したところ, 約 28 dB(A) であった。

3 結果

まず, 各刺激の弁別閾と標準偏差を表 1 に示す。合成音声 I と II の弁別閾は数詞の平均値である。調波複合音の弁別閾が最も小さく, さらに合成音声 II が合成音声 I よりも弁別閾は小さい。Tukey-Kramer の HSD 検定の結果, 5%水準で調波複合音, 合成音声 I, 合成音声 II のすべての組み合わせで, 有意な差が見られた。

次に数詞ごとの弁別閾を図 1, 2 に示す。合成音声 I は数詞によるばらつきが多い。一方, 合成音声 II は, 数詞に関係なく同程度の弁別閾の値である。

合成音声のデータについて, 数詞と基本周波数の二要因における分散分析を行ったところ, 数詞 (N), 基本周波数の変化 (F) の主効果, および交互作用効果 $N \times F$ がそれぞれ 5%水準で有意であった [$F(9,9) = 2.508, p = 0.0082$; $F(1,1) = 34.5105, p < .0001$; $F(9,9) = 2.4460, p = 0.0121$]。

合成音声 I のデータについて HSD 検定

表 1 各刺激の弁別閾の平均と標準偏差 (cent)

刺激の種類	平均	標準偏差
調波複合音	6.5	3.3
合成音声 I	21.2	10.3
合成音声 II	15.9	6.2

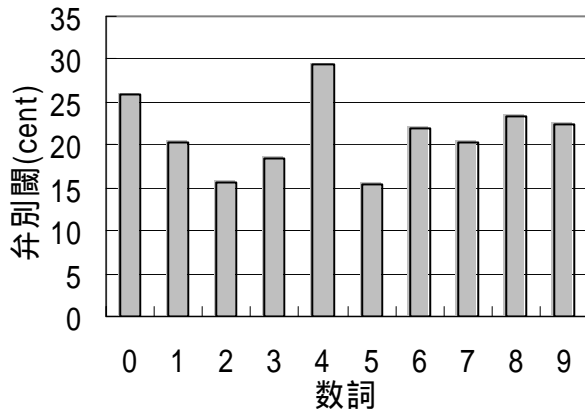


図 1 合成音声 I の各数詞の平均弁別閾：実験参加者 10 名の平均値。

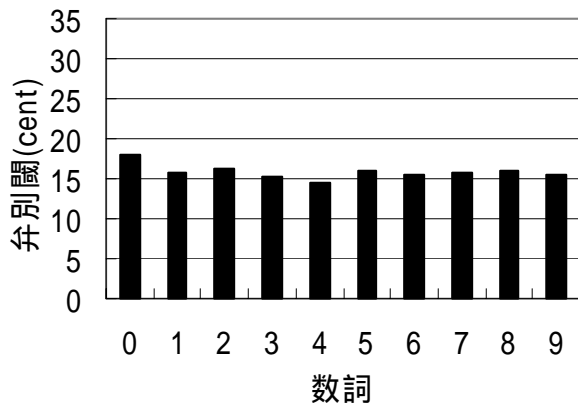


図 2 合成音声 II の各数詞の平均弁別閾：実験参加者 10 名の平均値。

を行ったところ、数詞「4」と「5」の間に 5%水準で有意な差がみられ、それ以外の数詞には差はみられなかった。合成音声 II のデータについて実験参加者の要因をブロック因子として数詞の要因の分散分析を行ったところ、実験参加者 (P) の主効果のみが有意であった [$F(9,81) = 32.4720, p < .0001$]。数詞 (N) の効果は有意ではなかった [$F(9,81) = 0.6236, p =$

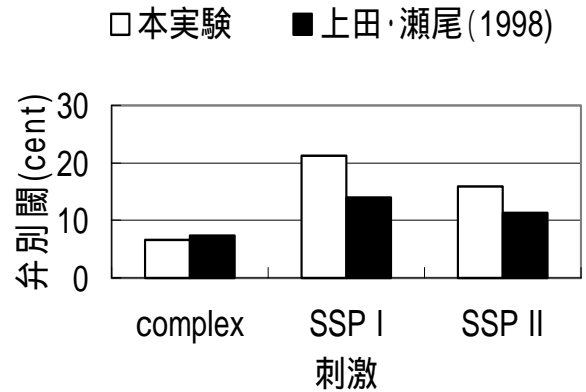


図 3 本実験と上田，瀬尾のデータ比較：SSP I は合成音声 I，SSP II は合成音声 II をあらわす。

0.7737]。また、各数詞の合成音声 I と合成音声 II について t 検定を行ったところ、「0 (ゼロ)」、「4 (よん)」、「9 (きゅう)」、「8 (はち)」が 5%水準で有意であった。

4 考察

4.1 上田，瀬尾 (1998) との比較

本実験で得られたデータと上田，瀬尾のデータを図 3 に示す。上田，瀬尾のデータと比較して、合成音声 I は 7.2 cent，合成音声 II は 4.6 cent 平均弁別閾が大きかった。また、すべての刺激において、上田，瀬尾のデータよりも標準偏差が大きかった。合成音声 I の弁別閾よりも、合成音声 IIの方が、有意に小さかったことから、基本周波数を定常にすることによって、音の高さが明確になったといえる。これは、上田，瀬尾の実験結果と一致する。上田，瀬尾の実験では、調波複合音と合成音声 II の間に有意な差はなかったが今回の実験では有意な差が見られた。

4.2 各刺激の弁別閾の考察

4.2.1 調波複合音

調波複合音の弁別閾は、6.54 cent であった。基本周波数が 110 Hz であるので、Hz に換算すると 0.42 Hz となる。これは、過去の研究結果 (Moore, 1973 など) と比

較しても十分に妥当なものであるといえる。本実験では、非音声を統制条件として用いた。従って、本実験で求めた弁別閾は信頼のあるものといえる。

4.2.2 合成音声 I

「2(に)」、「3(さん)」、「5(ご)」は、基本周波数が定常に近いので弁別閾が小さいと考える。一方、「0(ゼロ)」と「4(よん)」は、基本周波数の変化の幅が大きかったので弁別閾が大きくなったと考える。ではなぜ、数詞の中で最も大きな変化の幅を持つ「9」は「0」や「4」よりも弁別閾が大きくならなかったのか。これは、「9」は、単調に基本周波数が下降しているのに対し、「0」や「4」は、一度上昇してから、下降しているためと考える。

つまり、基本周波数の変化が弁別閾に及ぼす影響は、基本周波数の変化の幅が大きくても、変化が単調であれば、弁別閾はそれほど大きくなり、変化の幅が小さくても変化が単調でなければ、弁別閾は大きくなるといえる。

また、実験参加者の中でのばらつきが大きく、「2」や「5」などは、平均弁別閾が、1ステップ(12.5 cent)程度であった。従って、セント幅を小さくすると、数詞による差がより明確になると考える。

4.2.3 合成音声 II

各数詞の合成音声 I と合成音声 II について t 検定を行った結果から、基本周波数を定常にすることによって、基本周波数の変化が大きかった数詞の弁別閾は小さくなり、もともと基本周波数の変化が小さかった数詞の弁別閾は変わらないといえる。

しかし、基本周波数の変化がほとんどみられない「8」の合成音声 I の弁別閾は、合成音声 II の弁別閾よりも有意に高かった。この原因としては、振幅と基本周波数の関係が考えられる。つまり、振幅が大き

いときに、基本周波数の変化があるかどうかによって弁別閾が変わると考える。

5 まとめ

本研究では、合成音声の基本周波数の変化が音の高さの弁別閾にどのような影響を及ぼすのかを調べるために、音の高さの弁別閾の測定に関して上田、瀬尾の実験方法を改善し、より多くの実験参加者について実験を行った。そして、上田、瀬尾の実験結果と比較を行った。

その結果、次のことが明らかになった。

- (1) 合成音声の基本周波数を定常にすることで、弁別閾が有意に小さくなった。つまり、基本周波数を定常にすることが、音の高さの明確さにつながるということがわかった。これは、上田、瀬尾の実験結果と一致する。
- (2) 基本周波数の変化の幅が大きくても、変化が単調であれば、弁別閾はそれほど大きくなり、基本周波数の変化の幅が小さくても、変化が単調でなければ、弁別閾が大きくなる傾向がみられた。

文献

- Deutsch, D. (1970). "Tones and numbers: Specificity of interference in immediate memory," *Science*, **168**, 1604-1605.
- Kawahara, H., Katsuse, M. I., and Cheveigné, A. (1999). *Speech Communication*, **27** (3)-(4), 187-207.
- 勾坂芳典, 浦谷則好 (1992). "ATR 音声・言語データベース," *日本音響学会誌*, **48**, 878-882.
- Moore, B. C. J. (1973). "Frequency difference limens for short duration tones," *J. Acoust. Soc. Am.* **54**, 610-619.
- 上田和夫(2002). "音声と記憶", *心理学の方法*(ナカニシヤ出版), 141-158.
- 上田和夫, 瀬尾直子 (1998). "聴覚短期記憶の干渉効果: 音声の高さの明確さによる変化," *日本音響学会聴覚研究会資料*, H-98-9.