

7M-5

複数話者の識別に関する 微小波形変動の影響

越田一郎, 末広直樹, 伊藤憲治⁺, 田嶋隆一東京工科大学, ⁺東京大学

1. はじめに

人間は、複数の話者が同時に発生する音声を聞き分け、その中の1つだけに注意を集中して聞き取ることが出来る。この現象は、カクテルパーティー効果と呼ばれ、人間の音声認識上重要な課題となっている。このような識別を可能にする要因としては、1)音源方向、2)会話のコンテキスト、などが考えられる。しかし、これを可能にするためには、まず複数の音源が存在していることが認識されなければならない。これは、人間の音声認知システムの中でかなり下位の部分で行なわれていると考えることができる。

被験者に対して左右同じ刺激を与える、すなわち音源がモノラルである場合については、筆者らは報告している[1]が、ここでは、左右の刺激を異なるものとし、より識別しやすい場合の実験結果について述べる。

2. 実験方法

2.1 刺激材料

実験に用いた音声刺激は、無響室内で発声した単母音 /ア/、/エ/ (1音の長さは約 200ms, 平均基本周波数は約 200Hz) をサンプリング周波数 20kHz, 量子化ビット数 14 ビットでAD変換したものである。これを、次の様に加工して用いた。

・ <ア1><エ1>

無加工の音声から典型的な1ピッチ波形を切り出して、これを繰り返したものである。したがって、瞬時基本周波数の変動、各ピッチ波形の微小な違いは存在しない。

・ <ア2><エ2>

無加工の音声の各ピッチの瞬時基本周波数変動をなくし、基本周波数の 200Hz に固定したものの、これは、各波形の長さを伸縮して作製してあるので、各ピッチ波形の微小な違いは保たれている。

・ <ア3><エ3>

上記<ア1><エ1>に無加工の音声の自然な瞬時基本周波数変動を与えたもの。各ピッチ波形は同じであるが、ピッチ周期が変動している。

・ <ア4><エ4>

無加工の自然音声そのもの。

なお、以上の音声刺激は先頭波形での位相をそろえている。

2.2 実験内容

上記のような音声刺激および無音声刺激の中から、ランダムに2音をとりだし、聞こえた音が1音か2音を判断させるという実験を、以下の4種類の場合について行なった。

・セッション1

異なる音声刺激グループ<ア>、<エ>を用い、<ア>はLチャンネル、<エ>はRチャンネルに固定した。

・セッション2

音声刺激はセッション1と同じであるが、Lチャンネル、Rチャンネルの出力を固定にせず、ランダムに変動させた。

・セッション3

音声刺激グループ<ア>だけを用いて行なった。

・セッション4

セッション3と同じであるが、音声刺激グループ<エ>を用いた。

3. 実験結果

実験結果を表1から表4に示す。表中、'/ 'の左の数字は2音と判断した回数、右は1音と判断した回数である。被験者の性別および年齢は表5のとおりである。

4. 検討

・セッション1

<ア1>と<エ1>の場合では、ほとんどの被験者が1音であると判断している。また、<ア3>と<エ3>の場合も1音であると判断している場合が多い。これ以外の場合では、1音と判断している場合は少ない。このことから、複数話者の判定は、波形の時間変化の有無が大きな要因になっていることがわかる。また、上記2種類のうち、<ア1><エ1>の場合はほとんど1音であると判断していることを考えると、場合の数を比較してみると、瞬時基本周波数の変動も、副次的な要因となっていると言える。

・セッション2

基本的な傾向はセッション2と同じであるが、2音であると判断した場合がセッション1に比べると多くなっている。左右のチャンネルを固定にするか否か、だけがセッション1と2の違いであるから、波形の時間変化に対して左右の優位性が存在すると考えられる。

・セッション3, セッション4

この場合でも、<ア1>と<ア3>および<エ1>と<エ3>の組み合わせで1音と判断している場合が多い。したがって、波形の時間変化の影響は音韻が同じ場合でも変わらないものであると言える。

5. むすび

複数話者の判別に対して、

- 1) 波形の時間変化
- 2) 瞬時基本周波数の変動

が影響することを示した。

<参考文献>

- [1] 越田, 末広, 松本, 伊藤: 「人間の認知機能を考慮した情報論II」, 信学技報 IT87-92, 1988

表1 セッション1 の結果

R L	S	ア1	ア2	ア3	ア4	無音
	B					
エ1	U	0/5	5/0	3/2	5/0	0/5
	T	1/4	5/0	3/2	5/0	0/5
	K	0/5	5/0	4/1	5/0	0/5
エ2	U	5/0	5/0	5/0	5/0	0/5
	T	5/0	5/0	4/1	5/0	0/5
	K	5/0	4/1	5/0	5/0	0/5
エ3	U	5/0	5/0	3/2	5/0	1/4
	T	0/5	3/0	3/2	5/0	0/5
	K	5/0	4/1	3/1	5/0	0/5
エ4	U	5/0	5/0	5/0	5/0	0/5
	T	5/0	5/0	5/0	5/0	0/5
	K	5/0	5/0	5/0	5/0	0/5
無音	U	0/5	0/5	0/5	0/5	0/0
	T	0/5	0/5	0/5	0/5	0/0
	K	0/5	0/5	0/5	0/5	0/0

表2 セッション2 の結果

	S	ア1	ア2	ア3	ア4	無音
	B					
エ1	U	0/5	5/0	4/1	5/0	0/5
	T	0/5	5/0	4/1	5/0	0/5
	K	0/4	5/0	5/0	5/0	1/3
エ2	U	5/0	5/0	5/0	5/0	0/5
	T	5/0	5/0	5/0	5/0	0/5
	K	3/1	5/0	5/0	5/0	0/4
エ3	U	5/0	5/0	5/0	5/0	0/4
	T	4/1	5/0	3/2	5/0	0/5
	K	2/1	5/0	3/1	4/0	3/2
エ4	U	5/0	5/0	5/0	5/0	0/5
	T	5/0	5/0	5/0	5/0	0/5
	K	5/0	3/1	5/0	4/0	0/5
無音	U	0/5	0/5	0/5	0/5	0/0
	T	0/5	0/5	0/5	0/5	0/0
	K	0/5	0/5	0/4	0/4	0/0

表3 セッション3 の結果

R L	S	ア1	ア2	ア3	ア4	無音
	B					
ア1	U	1/4	5/0	1/2	5/0	0/5
	T	0/5	5/0	1/4	5/0	0/5
	K	1/4	4/1	4/0	5/0	0/5
ア2	U	3/0	0/4	5/0	4/0	0/5
	T	4/0	0/5	5/0	5/0	0/5
	K	5/0	1/3	4/0	5/0	0/5
ア3	U	1/4	4/0	1/3	5/0	0/5
	T	4/1	5/0	0/5	5/0	0/5
	K	5/0	4/1	1/4	4/0	0/5
ア4	U	5/0	5/0	5/0	0/5	0/5
	T	5/0	5/0	5/0	0/5	0/5
	K	5/0	4/1	4/0	1/4	0/5
無音	U	0/4	0/5	0/5	0/5	0/0
	T	0/5	0/5	0/5	0/5	0/0
	K	0/5	0/5	0/5	0/5	0/0

表4 セッション4 の結果

R L	S	エ1	エ2	エ3	エ4	無音
	B					
エ1	U	0/5	5/0	5/0	5/0	0/5
	T	0/5	5/0	4/1	4/1	0/5
	K	1/3	5/0	4/1	5/0	1/4
エ2	U	5/0	0/4	5/0	5/0	0/5
	T	5/0	0/5	5/0	5/0	0/5
	K	5/0	1/4	5/0	5/0	1/4
エ3	U	3/1	5/0	3/2	5/0	0/5
	T	2/3	5/0	0/5	5/0	0/5
	K	4/0	4/1	2/3	5/0	0/5
エ4	U	5/0	5/0	5/0	0/5	0/5
	T	5/0	4/1	5/0	0/5	0/5
	K	5/0	5/0	4/0	0/5	1/4
無音	U	0/5	0/5	0/5	0/5	0/0
	T	0/5	0/5	0/5	0/5	0/0
	K	0/5	0/5	0/5	1/4	0/0

表5 被験者

被験者	性別	年齢
U	M	23
T	M	24
K	F	22