

# 破裂子音のささやき音の特徴

東 恒人\*・古川 暢隆\*\*・高山 和敏\*\*\*・中川紀美雄\*\*\*\*

\*岡山理科大学工学部情報工学科

\*\*ベスト電器

\*\*\*岡山理科大学大学院工学研究科電子専攻

\*\*\*\*岡山理科大学工学部電子工学科

(1996年10月7日 受理)

## 1. まえがき

ささやき音を自動認識するようなシステムを構築できれば、ささやき音でしか日常の会話ができない人の助けとなろう。このようなシステムを実現するためには、ささやき音の母音および子音の特徴を明確にしておく必要がある。しかし、ささやき音に関する報告<sup>1)2)</sup>は有聲音<sup>3)4)</sup>に比べて少ないようである。

ささやき音は有聲音に比べて雑音性が強いため、子音については、ささやき音の波形構造、ホルマント周波数（以下、ホルマントという）ならびにささやき音の合成音の特徴が不明確である。本論文では、子音のうち破裂音<sup>3)~5)</sup>のささやき音の特徴を有聲音に対比させて以下の観点から論じる。①波形構造、②ホルマント、③合成音。

## 2. 計測

### 2.1 計測対象

計測対象は、破裂音<sup>5)</sup>に属するカ行とタ行の子音“カ, キ, ク, ケ, コ, タ, チ, ツ, テ, ト”（以下、これらを、ka, ki, ku, ke, ko, ta, ti, tu, te, to と表すが、誤解が生じない場合にはカ行の子音部の総称して k, タ行の子音部を総称して t で表す。）の有聲音とささやき音である。また、被験者は男子大学生 4 名（発声の訓練を受けていない）であり、同一の被験者に対して計測を 2 回づつ行った。

### 2.2 記号の定義

耳に聴こえる子音を Xy（ここでは、X=k, t, また, y=a, i, u, e, o を意味する）とすると、子音は子音らしさを表す部分（以下、主因子という）X と個々の子音の識別に寄与する後続母音部分<sup>5)</sup>（以下、副因子という）y からなっている。後述するように、計測した子音 Xy を主因子 X と副因子 y に完全に分離することは困難である。このような場合、主因子を X(y) で、副因子を (X)y で表す。さらに、有聲音の子音とささやき音の子音の

区別が必要な場合には、それを  $UX_y$  と  $Sxy$  のように記述する。

### 2.3 計測手順

計測に用いた装置類<sup>2)</sup>を表1に示す。これらの装置類を用いた計測の手順を以下に示す。なお、この手順は一部を除いて文献<sup>2)</sup>の手順と同様である。

- ① 標準マイクに音圧校正器（発振周波数 1 kHz, 出力音圧 1 Pa）をあて、DAT の録音レベルが -20dB となるように DAT の input level を調整する。
- ② ポップノイズを防ぐため、マイクロホンから 5 cm のところに防止膜を挿入し、また、マイクロホンから 10 cm のところから発声する。
- ③ 発聲音を標準マイクに入力し、その出力を DAT で記録する。
- ④ DAT からの再生音を、パーソナルコンピュータの AD 変換器ボードにより、16 kHz でサンプリングする。なお、量子化ビット数 16 で、低域フィルタの遮断周波数は 20 kHz である。
- ⑤ 10 msec の時間窓長のハニング窓を用いて、分析区間（波形）をサンプリングデータから切り出す。
- ⑥ 切り出した波形をパーソナルコンピュータの DA 変換器ボードに入力する。
- ⑦ DA 変換器ボードの出力を、アンプとスピーカを用い、再生音が所要の子音として認識できるかどうかを確認する。
- ⑧ 確認できた波形を分析次数 20 で LEC 分析し、ホルマントを求める。

なお、①～③の手順は暗騒音レベル 16 dB (A) の防音室内で実施された。

### 2.4 計測結果

#### 2.4.1 波形とソナグラフ

##### (1) 分析区間の決定

有聲音とささやき音のそれぞれの子音の計測波形の例を図1に示す。目的の音が 2.

表1 装置類

品 名	型 式	製 造 元
標準マイク	4134	B & K
標準マイク用アンプ	2803	B & K
音圧校正器	4230	B & K
DAT	DTR-100P	DENON
パーソナルコンピュータ	PC-9801EX4	NEC
拡張メモリボード	EMJ-4000L	メルコ
AD/DA 変換部	Sound Master	カノープス電子
アンプ	Integra A-917	ONKYO
スピーカ	101M	BOSE

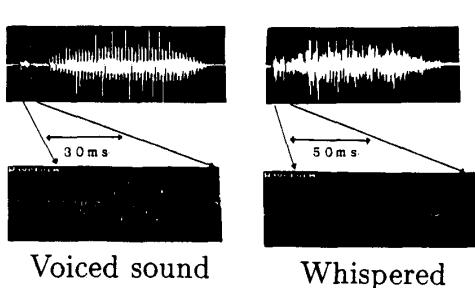


図1 子音の波形例

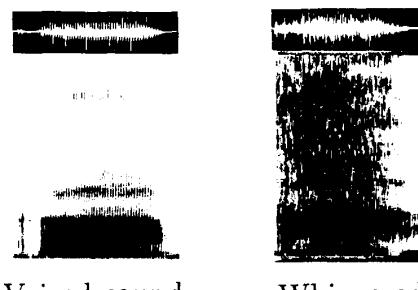


図2 子音の波形およびソナグラフの計測例

3の手順⑦により確認できるような波形部分を分析区間とした。すなわち、同図の枠内の部分が主因子波形に相当し、その継続時間は、有声音では約50ms、ささやき音では約30ms 前後である。そこで、主因子波形の分析区間を、有声音では約50ms、ささやき音では約30ms とした。また、有声音の副因子波形の分析区間として、発音が安定している任意の区間の50ms とした。後述するように、ささやき音の主因子波形と副因子波形との境界は不明確である。そこで、ささやき音の副因子波形の分析区間を、主因子の影響を避けるために、発声の終了近辺の50ms とした。

## (2) 主因子 $X(y)$

ささやき音の場合、有声音の場合に比べて全体的に雑音性が強いために、エネルギーの小さな主因子波形と副因子波形との境界は不明確である。

- ①  $X=k$  の場合の波形とソナグラフの計測例を図2に示す。有声音では主因子波形  $X(y)$  と副因子波形  $(X)y$  との境界はおおよそ推定できるが、ささやき音ではそのような境界を見いだすのは困難である。
- ② 有声音の  $ti$ ,  $tu$  は破擦子音<sup>5)</sup>に属する。このため  $ti$ ,  $tu$  は  $ta$ ,  $te$ ,  $to$  に比べて雑音性が強く、それらの主因子波形  $X(y)$  と副因子波形  $(X)y$  との境界は不明確である。 $X=t$  の場合のささやき音では、 $X=k$  の場合のささやき音と同様の傾向である。

## (3) 副因子 $(X)y$

有声音の副因子波形は明確な振動性波形であり、一方、ささやき音の副因子波形は雑音性波形である。これらの結果は短母音の結果<sup>3)</sup>と類似している。

### 2.4.2 フォルマント分布

$X=k$ ,  $t$  の副因子のフォルマントの平均値をそれぞれ表2, 3に、また、主因子のフォルマントの平均値をそれぞれ表4, 5に示す。表2～5を用い、フォルマント分布の例として第1フォルマント  $F_1$  と第2フォルマント  $F_2$  の関係を図3に示す。なお、図を見やすくするために、副因子  $(X)y$  を単に /y/ で表示している。また、有声音の場合を実線で、ささやき音の場合を破線で表している。

表2 (X) y のホルマント平均値 (X=k)

	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
(X) a	1001.93	1325.51	2965.62
	831.58	1281.59	2710.53
	766.92	1210.50	2952.63
(X) i	447.38	2463.16	3026.32
	318.42	2268.42	3081.57
	289.47	2274.70	2928.50
(X) u	688.23	1637.65	2856.32
	343.42	1659.21	2809.50
	319.54	1251.90	2266.90
(X) e	793.86	2206.14	2993.33
	519.74	2005.22	3058.20
	507.51	1951.10	1575.10
(X) o	785.32	1341.27	3114.96
	560.53	918.42	2986.85
	537.59	921.05	2575.10

上段：ささやき声

中段：有聲音

下段：有聲音（短母音）<sup>2)</sup>

表3 (X) y のホルマント平均値 (X=t)

	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
(X) a	931.50	1815.79	3121.05
	876.55	1489.43	2994.73
(X) i	589.47	2005.26	2794.74
	363.27	2163.15	3194.74
(X) u	763.16	1647.37	2815.81
	373.69	1689.47	3310.53
(X) e	773.70	2089.47	2826.31
	531.50	2005.26	2763.16
(X) o	805.27	1184.20	2942.10
	605.26	944.74	2968.16

上段：ささやき声

下段：有聲音

これらの表および図から以下のことがわかる。

## (1) 副因子 (X)y

有聲音とささやき音の各場合とも、副因子 (X)y のフォルマントは短母音の結果<sup>2)</sup>よりも高めの数値が得られているが、全体としての傾向は短母音の結果<sup>2)</sup>と類似している。すなわち、副因子 (X)y は短母音<sup>2)</sup>に類似した五角形分布する。また、ささやき音の五

表4  $X(y)$  のホルマント平均値 ( $X=k$ )

	$F_1$	$F_2$	$F_3$
$X(a)$	953.20	1601.97	3251.97
	1430.95	3016.16	4266.45
$X(i)$	796.06	1763.16	3197.37
	1286.14	2566.47	3557.89
$X(u)$	747.37	1840.87	3347.36
	1192.12	2471.35	3728.65
$X(e)$	785.09	2276.32	3167.27
	1206.77	2364.66	3642.86
$X(o)$	828.95	2175.51	3232.45
	1008.77	2372.69	3736.84

上段：ささやき声

下段：有聲音

表5  $X(y)$  のホルマント平均値 ( $X=t$ )

	$F_1$	$F_2$	$F_3$
$X(a)$	931.58	1815.79	3121.05
	131.58	921.05	1710.10
$X(i)$	1210.53	2118.42	3065.79
	858.02	2036.84	3173.68
$X(u)$	1175.44	2342.31	3640.35
	594.74	1626.32	2794.74
$X(e)$	847.37	2110.53	2952.70
	157.89	1315.79	2263.16
$X(o)$	868.42	1526.32	2960.68
	342.11	1289.47	3026.32

上段：ささやき声

下段：有聲音

角形分布は有聲音の五角形分布よりも高周波側にずれる傾向にある。このことから、副因子 ( $X$ )  $y$  に及ぼす主因子  $X(y)$  の影響は少ないといえる。

## (2) 主因子 $X(y)$

① 有聲音の場合、主因子と副因子の各々のフォルマント分布は異なっているが、ささやき音の場合、主因子と副因子の各々のフォルマントが一部重なった分布となつておらず、ささやき音の主因子  $X(y)$  のフォルマントは副因子  $y$  のフォルマントの影響を受けているといえる。

② ささやき音のフォルマントは有聲音のフォルマントよりも  $X=k$  の場合には低く、 $X=t$  の場合には高くなるように分布する。以上のような特性の主因子と副因子を用

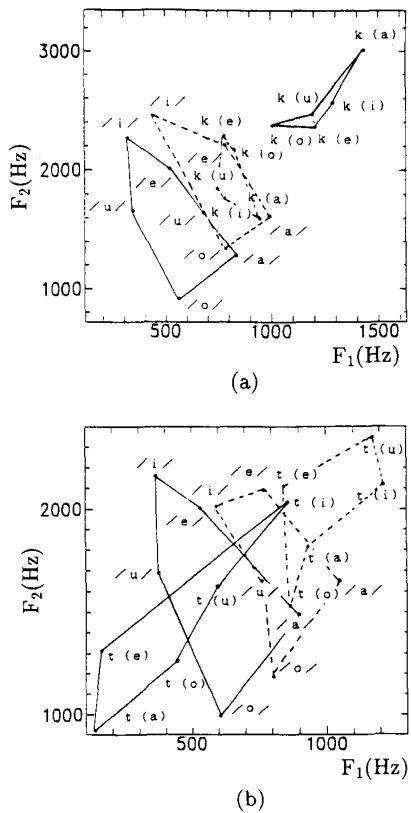
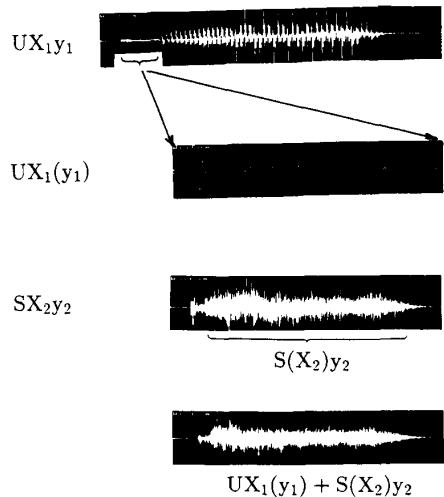
図3  $F_1$ と $F_2$ の関係

図4 波形合成の手順

いて、音声を合成し、その合成音の特徴を調べる。

### 3. 合成音

#### 3.1 記号の定義

すでに計測した子音  $Xy$  と  $Wz$  を用いて音声合成する場合、子音  $Xy$ ,  $Wz$  を主因子  $X(y)$ ,  $W(z)$  と副因子  $(X)y$ ,  $(W)z$  に分解し、子音  $Xy$  の主因子  $X(y)$  と子音  $Wz$  の副因子  $(W)z$  との合成を  $X(y) + (W)z$  と記述する。

#### 3.2 合成手順と合成パターン

波形合成の手順を図4に示す。具体的には、発声した子音の主因子  $X(y)$  とこれとは別の子音（同一人が発声した別の種類の子音あるいは他人が発声した同一子音）の副因子  $W(z)$  を合成し、目的の子音  $Xz$  が認識できるかどうかを調べる。合成パターンを以下の2種類とした。

##### (1) 合成パターン1

同じ子音  $Xy$  ではあるが、異なる個性が発声した同一子音  $X_1y_1$ ,  $X_1y_2$  を分解し、 $X_1y_2$ ,  $X_2y_1$  を合成する。

## (2) 合成パターン 2

同一人が発声した異なる子音  $X_y, W_z$  を分解し,  $X_z, W_y$  を合成する。

## 3.3 合成音の特徴

各パターンの合成音の特徴は以下の通りである。

## 3.3.1 合成パターン 1

- ① 有聲音の主因子  $UX_1(y_1)$  と有聲音の副因子  $U(X_2)y_2$  との合成音は有聲音の子音として認識できる。

$$UX_1(y_1) + U(X_2)y_2 = UXy_2$$

- ② 有聲音の主因子  $UX_1(y_1)$  とささやき音の副因子  $S(X_2)y_2$  との合成音はささやき音の子音として認識できる。

$$UX_1(y_1) + S(X_2)y_2 = SXy_2$$

- ③ ささやき音の主因子  $SX_1(y_1)$  とささやき音の副因子  $S(X_2)y_2$  との合成音はささやき音の子音として認識できる。

$$SX_1(y_1) + S(X_2)y_2 = SXy_2$$

- ④ ささやき音の主因子  $SX_1(y_1)$  と有聲音の副因子  $U(X_2)y_2$  との合成音は有聲音の子音として認識できる。

$$SX_1(y_1) + U(X_2)y_2 = UXy_2$$

- ⑤ 有聲音およびささやき音のそれぞれの合成音の個性には、合成に用いた副因子の個性が現れる。

- ⑥ 有聲音の合成音にはささやき音の合成音よりも個性が強く現れる。

以上のことから、副因子が有聲音の成分あるいはささやき音の成分かに応じて、合成音が有聲音の子音あるいはささやき音の子音として認識できる。

## 3.3.2 合成パターン 2

- ①  $X(a)$  と  $(X)i$  を合成した場合、 $X_i$  として認識できるが、厳密には、 $X_{ai}$  として認識できる。これは  $X(a)$  の中に副因子の  $a$  の情報が含まれて、 $X_a$  となっているためである。この副因子の  $a$  の情報を可能な限り除去するため  $X(a)$  の波形から発声の始まりに向けて波形を削除する。正常な子音の継続時間長を保つために、継続時間が短くなった波形の後に空白時間を挿入する。この波形に副因子波形を接続しても、その合成音は副因子に応じた母音にしか認識できない。上記の内容を式で表現すると次式となる。

$$\begin{aligned} X(a) + a &= Xa \\ X(a) + u &= X_a u \doteq Xu \\ X(a) + o &= X_a o \doteq Xo \end{aligned}$$

②  $X(a)$  と  $(X)i$  あるいは  $(X)e$  との合成音は詰まった感じの短母音  $i$  あるいは  $e$  として認識される。

$$\begin{aligned} X(a) + i &= "i \\ X(a) + e &= "e \end{aligned}$$

このパターンの合成音は目的通りの子音として認識できない。この結果は主因子に副因子の成分が混入しているために生じたのであろう。

以上のことから、主因子と副因子が 2.4 に示した特性であっても、合成パターン 1 の合成音は子音らしく認識できることがわかる。

#### 4. む す び

本論文では、子音のうち破裂音のささやき音の特徴を有聲音と対比して調べた。その結果を以下に列挙する。① ささやき音の波形は雑音性が強く、主因子波形  $X$  と副因子波形  $y$  との境界は有聲音の波形ほど明確でない。② 副因子  $(X)y$  のフォルマント分布は短母音の結果<sup>3)</sup>と類似している。一方、主因子  $X(y)$  のフォルマントは副因子  $y$  の影響が混入しており、この影響を除くことは難しい。③ この程度の副因子  $y$  の影響が混入していても、個性が異なった同一子音  $Xy$  間の合成音でも子音として認識できる。④ 個性が異なった子音  $Xy$  間の合成音は、副因子  $y$  が有聲音の成分かささやき音の成分かに、また、 $a, i, u, e, o$  かに依存した子音として認識できる。⑤ 個性が同じでも子音  $Xy$  と  $Wz$  の主因子と副因子間の合成音は不完全な子音としか認識できない。

以上の結果から、破裂音の子音の特徴ならびに主因子に含まれる副因子の影響の除去は困難であるが、ある程度の音声合成は可能であることがわかった。今後は、他の種類の子音の特徴分析を進め、自動認識システムの実現の可能性について検討する必要がある。

#### 参 考 文 献

- 1) 小橋 豊：“音と音波”，基礎物理学選書 4，裳華房，p.170 (1976).
- 2) 東、平田：“短母音のささやき音のホルマント特性”，信学論(A), Vol.J77-A, No.9, p.1300-1303(1994).
- 3) 今堀克己：“音響分析”，理化集報 5，小山書店 (1949).
- 4) 日本音響学会編：“音声”，音響工学講座⑦，コロナ社 (1977).
- 5) 日本音響学会編：“音響用語辞典”，コロナ社 (1990).

## Characteristics of Whispered of Plosive Consonant

Tsunehito HIGASHI, Masataka FURUKAWA,  
Kazutoshi TAKAYAMA and Kimio NAKAGAWA

*Faculty of Engineering,  
Okayama University of Science,  
Ridai-cho 1-1, Okayama 700, Japan*

(Received October 7, 1996)

This paper describes characteristics of whispered by compearing with voiced sound in case of plosive consonant.

The following characteristics are shown.

- ① Formants of back vowel almost agree with those of isolated vowel.
- ② Back vowels influence on formants of constnant and attributes of synthesis sounds.