

# 東北の話者における有声性に関わる パラメータの探索とその世代差

高 田 三 枝 子

## 1. 研究の背景

日本語の語頭閉鎖音の VOT 値は地域差および世代差があり、特に若い世代ではプラス値の領域で有声音と無声音の重なりが生じている。このことから VOT 以外の何らかの音響パラメータが有声性に関与していることが予想される。本稿では、高年層も有声音の VOT がプラス値に分布する東北に注目し、高年層と若年層の 2 世代の話者の語頭破裂音の音声を資料として、いくつかの音響的特徴についての測定結果を報告し、またその測定値を一般化線形混合モデルに当てはめた時の当てはまりの良さの指標 (AIC) やデータの予測確率から有声性に関与する音響パラメータを検討する。

日本語の有声性の別に関わる音響パラメータは、従来様々なものが提案されてきた。Shimizu (1996) はアジアの 6 言語および英語について、閉鎖音系列の音韻的対立とその音響的特徴についての産出音声の音響分析をもとに通言語的に記述している。その結果、日本語の語頭閉鎖音に関しては、VOT、F0 開始周波数と遷移パターン、閉鎖開放から 25 msec のスペクトル特性、F1 開始周波数の関与を指摘した。また最近では、Tamura et al. (2018) が、ノイズ部分の amplitude と開始 F1 周波数の操作を施した /d/-/t/ の刺激音を用いた聴取実験を行い、その結果、F1 周波数の効果は見られないことと、一方でノイズ部分の amplitude の効果が見られ、これが大きい刺激で有声性弁別の VOT 境界が小さくなることを指摘した。

このように様々な音響パラメータが関与する可能性がある中で、特に中心的な音響パラメータと

して有効性が注目されてきたのが VOT である。しかしこの VOT については近年世代間で急激に変化し、若い世代では有声音と無声音の間での値の重なりが指摘されている (高田 2011)。これに伴い、VOT 以外の音響パラメータの重要性が増しているのではないかという推測から、近年、複数の研究が示され、特にピッチの重要性が注目されるようになってきている。Gao & Arai (2019) と邊 (2019) はどちらも、若い世代の音声を分析し、VOT に加えピッチ (f0) が有声性の違いに大きく関与していることを指摘している。さらに邊 (2019) は日本国内の 4 地域 (東北、中部、近畿、九州) について比較しており、VOT と f0 という二つのパラメータ (あるいはどちらか) で大方有声性の判別がつくこと、ただしどちらのパラメータがどれほど区別に関与するかは地域によって異なることを指摘している。このうち東北については他地域に比べピッチの関与が大きいことを指摘した。

ただし、Gao & Arai (2019) も邊 (2019) も若い世代のデータを扱っており、世代差については扱っていない。VOT の世代差を考えれば他の音響パラメータの関与についても世代差の存在が推測される。本稿ではこの世代差に着目したい。

なおこれまでに指摘された有声性に関わる音響的特徴は多数あり、理想的にはすべての音響パラメータの検証を視野に入れることが望まれるが、本稿はそのすべてを扱うものではない。本稿では途中段階の報告として、VOT、ピッチを中心として、ここに試みとして、破裂ノイズに関わる指標と、さらに声帯振動開始に対する母音の共鳴エネルギー (第 1・第 2 フォルマント (F<sub>1</sub>・F<sub>2</sub>)) の

遅れという現象を加えて検証した結果を報告する。共鳴エネルギーの遅れは、筆者が分析の過程で気になり、試みとして加えたもので、調音上これが何を意味するのか現時点で確固とした理解があるわけではない。その点、これをとりあげる根拠は薄い。しかし実際のところ、本稿の検証の結果これが有声性に一定度の関与を示す結果が得られたため、ここであえて提示するものである。

ところで、本稿では日本語の中でも東北という地域に注目する。東北は、語頭有声閉鎖音のVOTに関して、他方言と違い、古くから有声音のVOTがプラス値を取っていた地域である(高田 2011)。少なくとも1986年頃の調査の高年齢層(1900年前後生まれの話者を含む)において有声音のVOTが統一的にプラスの値を取っていた。また邊(2019)の若年層の音声の分析で、東北は他地域に比べ特にピッチの関与が大きいことが指摘されている。すなわち、もしこの地域において有声性の区別に関するモデルとして高齢層、若年層の違いがなく、ピッチの関与の大きいモデルが採用されるとすれば、他地域の若年層にも広がりを見せるピッチの関与がどこに端を発するものであるかの手掛かりとなる可能性がある。

## 2. 研究方法

本節では分析に用いる音声資料の話者、発話項目、分析項目(音響パラメータ)等、研究方法について説明する。

### 2.1. 資料

本稿で使用する資料は、筆者が2006～2007年に収集した録音調査資料である。話者は秋田県に在住、生え抜きの話者で、1920年代生まれ(当時78～85歳)の9名(男性5名、女性4名)と、1980年代生まれ(当時17～20歳)の10名(男性5名、女性5名)の2世代の計19名である。無意味語および有意味語を単語リストの形で提示、全て単独で(フレームに入れず)読み上げてもらった。方言的特徴が出にくいスタイルではあるが、話者の意識的コントロールの届かない有声性に関する音響的特徴を見る上では、その地域的な

特徴を見ることは可能であると考えられる。本稿で分析する語は、収録語のうち、次の無意味語である。

1 音節語：/pa, ta, ka, ba, da, ga/, 各3回

2 音節語：/tada, data/, 各12回(アクセント＝1型)

### 2.2. 各音響パラメータの測定方法

分析にあたり、各発話について、Praat (version 6.0.28) を使用して次の①～④のイベントをマークした。なお( )内の文字列はPraat上で付した記号である。②については、聴覚的な開始ではなく、 $F_1 \cdot F_2$ がそろって観察できるようなる始点である。また④については、東北の話者では声帯振動が破裂前に見られない場合がほとんどで、その場合にはvwStrより後の時点で仮にこれをマークした(大きなプラス値となるため、後で測定値から容易にはじくことができる)。

① 閉鎖解放時点 (bst)

② 閉鎖解放後の声帯振動開始時点 (pstv)

③ 母音  $F_1 \cdot F_2$  開始時点 (vwStr)

④ 閉鎖解放前の声帯振動開始時点 (prevv)

これらの具体的なマーク位置の例を図1に示す。

以上のイベントマークを利用して、本稿で注目する音声特徴に対応する音響パラメータを測定した。表1にその具体的な内容を示す。表中、音響パラメータの欄の( )内は各音響パラメータの省略形として以降、文中で用いる。

VOT測定の詳細については説明を省く(詳細は高田(2011)など)。pitchについては、この値が直接基本周波数(fo)でなく、第1倍音( $H_1$ )で代用されていることに注意が必要である。これはPraatでの測定において音声開始部のfoの測定に計測ミスが多発する問題に対するためである。intensityGap(以降、iGap)は外破のエネルギーの強さを母音開始時点を基準として相対的に捉えようとしたものである。これはTamura et al.(2018)で示された破裂のノイズの弁別への影響から加えたものである。ただし、測定範囲など、さらに検討の余地がある。またvDelayについては、

東北の話者における有声性に関わるパラメータの探索とその世代差 (高田)

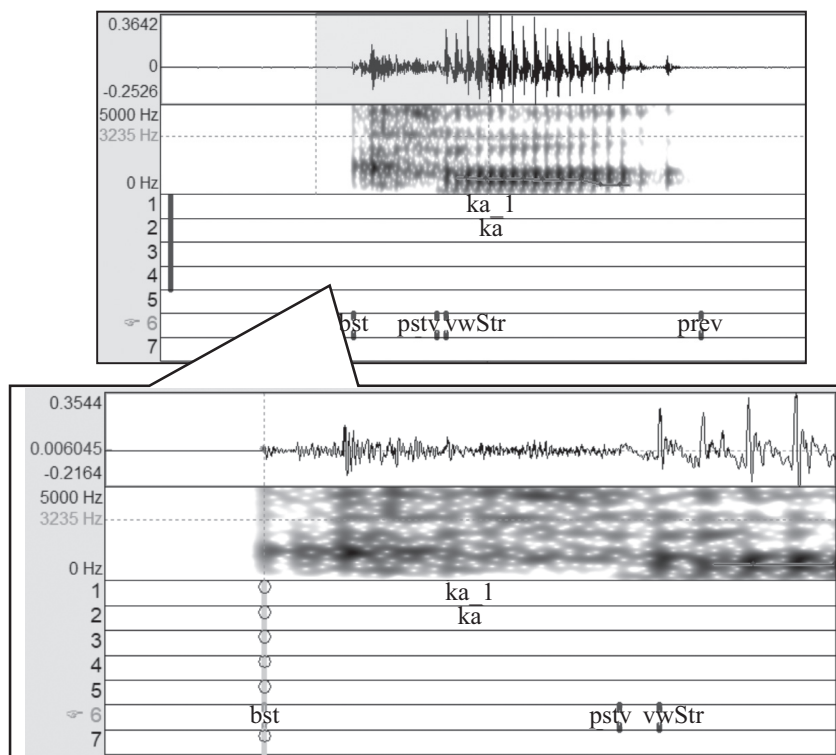


図1 Praat 上でのイベントマーキング具体例

表1 観察する音声特徴とその音響パラメータの測定方法

観察する音声特徴	音響パラメータ	具体的測定内容
声帯振動開始のタイミング	VOT	$A = bst(t) - prev(t)$ $B = pstv(t) - bst(t)$ VOT = A, B いずれか小さい方の値
声帯振動開始時の声の高さ	ピッチ (pitch)	pstv から 25ms の平均スペクトルにおける第 1 倍音周波数を mel に変換した後、各話者ごとに平均値と標準偏差で標準化
破裂のエネルギーの相対的強さ	母音開始時に対する外破の平均インテンシティー (intensityGap (iGap))	$A = bst$ から $pstv$ までの平均 intensity $B = vwStr$ から 25ms の平均 intensity $intensityGap = A - B$
声帯振動開始に対する母音開始の遅れ	声帯振動を表わすエネルギー開始に対する $F_1 \cdot F_2$ 開始の遅れ (vDelay)	$vDelay = vwStr(t) - pstv(t)$

先にも述べたように、その解釈が現時点で不十分で今後の検討を要する。ただ結果を先取りすれば、この音響パラメータは高年層においてはモデルに組み込むことで結果的にモデルの良さが上がることが確認され、「意味のある」要素であると思われる。これが調音上のどのような機序で生じ、また認知上利用されているのかということについて今後改めて考える必要がある。

### 3. 結果

以下ではまず各パラメータに関する単純な数値の分布について結果を示し(3.1節)、その後これらのパラメータを一般化線形混合モデルに当てはめた場合の当てはまりの良さ(AIC)と正判別確率からさらに考察を進める(3.2節)。

#### 3.1. 各パラメータにおける分布

図2～図5に各パラメータの測定値の分布を世代別に示す。各図とも箱ひげ図の箱部が色付きのものが有声音、白色のものが無声音を表わす。図2のVOTに関しては高年層・若年層とも一貫した傾向が見られる。すなわち有声音より無声音の方が大きい値に分布し、その分布域は多少重なりがあるものの限定的で、有声音の第三四分位数が無声音の第一四分位数を超えない(箱ひげ図の箱部分が重ならない)範囲にとどまる。この事から見ると、東北の話者における有声音と無声音はVOTによって、大部分において区別されることがわかる。

次に図3のpitchの分布を見てみよう。この値は測定値をmel値に変換した後、個人ごとに標準化した値であり、したがって図2で示した分布は個人差を捨象したものと考えることができる。これを見ると、高年層、若年層共に有声音より無声音の方が声帯振動開始時点のピッチが高い傾向を見ることができる。ただし、有声性による差の大きさについては世代差が見られ、若年層では有声音と無声音のピッチ差が大きい、高年層では差は小さく重なりも大きいことが見て取れる。

図4はiGapの値の分布である。iGapの値は結果として、全てプラスの値であった。このことは、

母音開始部の方が外破部の平均intensityよりも一貫してエネルギーが強いことを意味する。ただし、有声音と無声音との間での値の大小関係としては、若年層では一貫して無声音の方が値は大きくなっているが、高年層では軟口蓋音のみ、その他と逆の傾向(有声音の方が値が大きい)を見せており、一貫していない。なお、無声音の方がiGapの値が大きいということは、母音の開始時点に対する外破の平均intensityが、有声音よりも相対的に弱いということである。そのような結果は、Tamura et al. (2018)の結果から推測する破裂のノイズの在り方とは食い違う。このことには、今回のパラメータの測定において外破全体の平均値を用いたことが影響した可能性がある。つまり外破が長いほど平均値としては小さくなった可能性があり、これは一種の破裂のノイズの指標であるとはいえ、単純に破裂の「強さ」としては解釈できず、注意が必要である。破裂の強さを表すパラメータの測定方法については再考する必要がある。

図5はvDelayすなわち声帯振動開始に対する母音開始の遅れの値の分布である。これも先のiGapと同様、若年層では一貫して無声音の方が大きくなる傾向を見せるが、高年層では軟口蓋音のみ有声音の方が大きくなっており、一貫していない。なお、具体的な音声に戻ってみると、このvDelayの値が大きくなる場合F<sub>1</sub>よりF<sub>2</sub>が遅れる(はつきり出現するまでに)ことによるものが多く見られる印象がある。当初本稿でこのF<sub>1</sub>・F<sub>2</sub>開始の遅れは母音開始の遅れを表わす指標としてとりあげたが、聴覚的には充分母音が聴取できる時点となる場合もあり、このパラメータの解釈も再検討する余地がある。

以上の結果を踏まえると、ここで扱った四つのパラメータについて単純に考えれば、VOTとpitchは有声性の違いに大きく関与するが、それに比べiGapとvDelayはあまり関与しないのではないかと推測される。しかしこのことについて、一般化線形混合モデルによる結果を見ると、そう単純なことではないことがわかる。次節以降で詳しく見ていくことにする。

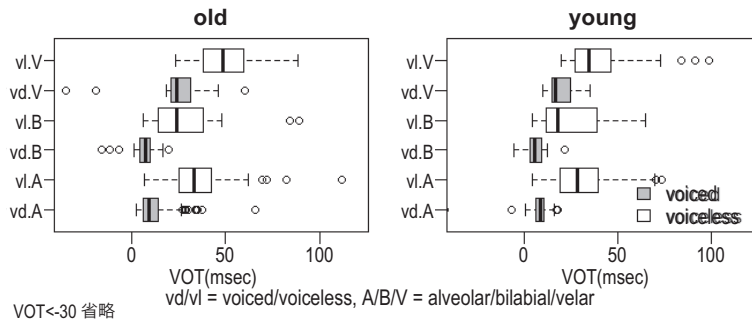


図2 VOT 値の分布

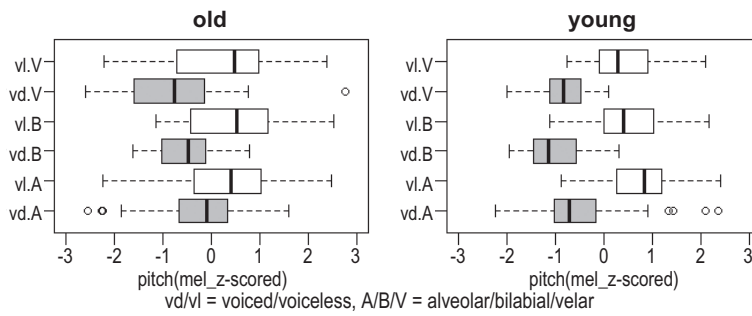


図3 声帯振動開始時点のピッチ (H1 値) の分布

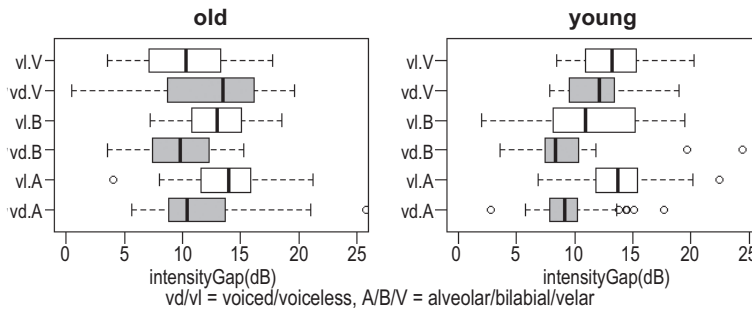


図4 外破と母音開始時点の intensity 差の分布

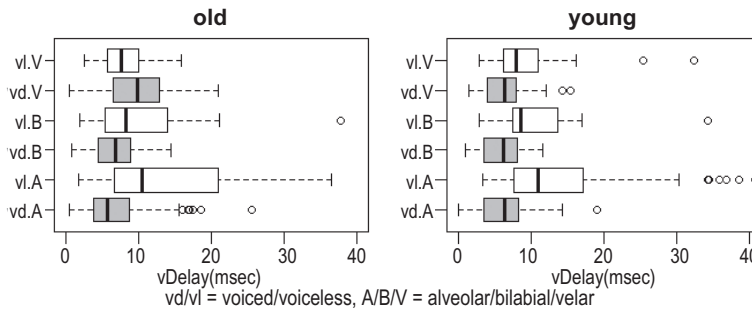


図5 声帯振動開始に対する母音フォルマント開始の遅れの分布

### 3.2. 一般化線形混合モデルによる検証

本節では一般化線形混合モデルを用いて各パラメータ投入の妥当性を検証する。その際、モデルの複雑さ（パラメータの数）を考慮に入れた統計モデルの良さを評価する指標である Akaike's Information Criterion (AIC) の値を利用する。AIC はこれが小さいほど良いモデルであるとされる。本節ではまず各パラメータを組み込んだモデルの AIC を比較し (3.2.1 節)、さらに各モデルをデータに当てはめ、個別データの予測確立（無声音である確率）を散布図で示し、詳細を見ることにする。

#### 3.2.1. 各パラメータの投入とモデルの AIC

モデルは有声・無声の 2 項分布とし、計算には R (ver.3.6.1) の “lmerTest” パッケージの glmer 関数を利用する（リンク関数は binomial）。モデルの設定にあたり、本稿では世代によって有声性に関わるパラメータが変化している可能性を考えると、世代別にモデルを仮定する。両世代のモデルとも、話者による変動をランダム効果として切片に設定する。また、調音位置は固定効果として常に含めるものとする。検討する音響パラメータは、先の 3.1 節と同様、VOT, pitch, iGap, vDelay の 4 つのパラメータである<sup>1)</sup>。係数（影響度）比較のため、全変数に対し各変数について総平均（全話者全発話の平均）と標準偏差により正規化を行った上でモデルに投入した（標準化は AIC には影響しない）。

まず各パラメータを単独で入れた場合の AIC を見たところ、表 2 のようになった。これを見ると、両世代とも VOT によるモデルの AIC が他のパラメータに比べ圧倒的に小さく、このパラメータの影響が大きいことが見受けられる。若年層においては pitch によるモデルもそれ以外に比べ AIC 値が小さいが、それでも VOT による方がさらに大幅に小さい。この点は邊 (2019) で東北の若年層の分析から VOT と pitch が同程度あるいは pitch の方が判別精度が高くなることもあるという結果とは異なる<sup>2)</sup>。

表 2 各パラメータ単独モデルの AIC

	高年層	若年層
VOT	276.81	215.21
pitch	490.16	310.8
iGap	508.39	458.13
vDelay	480.36	463.41

上の結果から、どちらの世代においても VOT はモデルに必ず入れるべき（産出面における有声性の区別と関係が深い）と考えられる。したがって以下、この VOT をベースとして、他のパラメータを一つずつ加えながら、モデルの AIC の動向をみることにする。

2 つ目のパラメータとして VOT 以外の、pitch, iGap, vDelay のどれが最も AIC を下げるか、また 3 つ目、4 つ目のパラメータを入れることで有意な AIC の下げ幅を得られるか、ということに注意して、順にパラメータを追加した。その結果を示したのが図 6 である。

図 6 上部の P は追加するパラメータの数および投入した順番を表している。各パラメータの横には AIC を [ ] 内に示し、さらに、それを投入する前のモデルに対して有意に AIC が小さいものについて右側に Signif. codes を示した。例えば高年層の P=2 の一段目は「vDelay[230.35]\*\*\*」となっているが、このモデルは VOT と vDelay の二つのパラメータを入れたモデルの AIC が 230.35 で、VOT のみを入れたモデルよりも有意に ( $p < 0.001$ ) AIC が小さいことを表している。図の描き方としては、各段階で、候補となるパラメータを入れたモデルの AIC を比較し小さいものから順に上から配置した。例えばパラメータを二つ入れるモデル (P=2) としては、VOT+vDelay の他に VOT+iGap, VOT+pitch があり得るが、高年層の場合この順に AIC が大きくなっていくのでこの順に並べて描いている。なお、複数のパラメータを含むモデルの場合、同じパラメータの組み合わせであれば、パラメータの投入の順番はその AIC に関わらない。例えば VOT+vDelay+iGap と iGap+vDelay+VOT は変わらない。

図 6 の若年層をみると、pitch が AIC の大幅な減

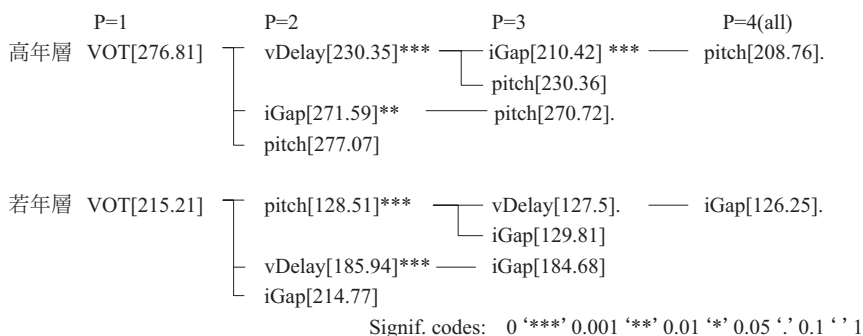


図6 各世代におけるモデルに投入するパラメータと各モデルのAIC

少をもたらし、他のパラメータ (vDelay と iGap) の投入は AIC を pitch ほど下げない。一方高年層では、音響パラメータ 2 つを投入するモデルとしては、VOT+vDelay のモデルが (VOT のみのモデルに対して) 最も大幅で有意に AIC が減少する。さらに 3 つのパラメータのモデルの場合はそこに iGap を加えた場合に有意に ( $p < 0.01$ ) AIC が減少する。Pitch はどの段階で加えても、AIC の下げ幅は小さく、有意差も他のパラメータと同じ水準 ( $p < 0.01$ ) では見られない。

つまり有声性を予測するのに最も良いと考えられるモデルが世代間で異なる。このことから VOT 以外の有声性に関わる音響パラメータが世代によって異なる可以考虑することができる。

ところで特に高年層の結果は、先の 3.1 節で各音響パラメータの測定値の単純な分布からは予測されづらい結果である。vDelay および iGap は高年層において調音点別に見た時に、軟口蓋音で他と異なる傾向を見せていた。このことから高年層の vDelay および iGap についてはさらに調音位

置 (place) との交互作用が存在する可能性が考えられる。そこで再度、これらの交互作用を想定した (主効果も含める) モデルを考えることにする。表 3 に調音位置と vDelay および iGap との交互作用なしとした場合とありとした場合の各モデルの AIC を示す。なお若年層についても高年層と同様に交互作用の有無で AIC を比したが、交互作用想定の一必要性は確認されなかったのでここでは省略する。

表 3 の AIC にみる通り、高年層は vDelay および iGap について調音位置との交互作用を想定した方が有意に AIC が小さくなる。したがって調音位置との交互作用は想定すべきであると考えられる。

さて次に本稿で検討した音響パラメータの相対的な影響の度合いを見ておきたい。4 つの音響パラメータすべてを投入したモデルにおける (高年層については iGap および vDelay の調音位置との交互作用も想定) 各パラメータの係数を表 4 に示す。なおモデルは無声歯蓋音を予測するモデルと

表 3 高年層における調査位置と vDelay, iGap との交互作用想定による AIC の相違

投入音響パラメータ	交互作用なし	交互作用あり
VOT+vDelay	[230.35]	[212.97]***
VOT+iGap	[271.59]	[263.68]**
VOT+vDelay+iGap	[210.42]	[183.69]***
VOT+vDelay+iGap+pitch <sup>※2</sup>	[208.76]	[179.52]*

※ 交互作用なしのモデルに対する有意差 (0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1)

※2 pitch を含めるモデルはモデルが収束しない警告が出る。

表4 モデル各パラメータの固定効果

高年層			若年層		
	係数	有意性		係数	有意性
切片	0.1411		切片	2.7093	**
VOT	9.0698	***	VOT	14.3558	***
vDelay	3.6288	***	vDelay	1.4603	.
iGap	-1.7476	***	iGap	-0.9027	.
pitch	0.7584	*	pitch	2.7363	***
調音位置 (両唇)	2.9437	***	調音位置 (両唇)	2.6140	***
調音位置 (軟口蓋)	-5.1965	***	調音位置 (軟口蓋)	-6.1596	***
vDelay×調音位置 (両唇)	-1.6613				
vDelay×調音位置 (軟口蓋)	-3.6618	***			
iGap×調音位置 (両唇)	2.5447	**			
iGap×調音位置 (軟口蓋)	-0.8308				

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

なっているため調音位置の係数として歯茎は表示されない。

表4を含め、以上の結果から考えれば、若年層の場合、有声性に関してVOTとpitchのみで十分説明がつくが、これに対して高年層ではむしろピッチの関与は小さく、他の多様なパラメータが必要になり、そのパラメータ間の関係性も複雑であるということが読み取れる。こうした点をより詳しく検討するため、次節では各モデルを実際のデータに当てはめ、各データごとに得られる予測確立と、実際の有声性との関係を見ていくことにする。

### 3.2.2. 各モデルのデータへの当てはめ

それぞれのモデルの詳細を見るため、各モデルを実際のデータに当てはめ、各データの予測確率（ここでは無声音である確率）を求めた。この予測確率（縦軸）とVOT（横軸）の値から散布図に描いたのが図7～図10である。

図7と図8はVOTともう一つ、計2つのパラメータによる3つのモデルを比較したものである。VOTのみのモデルも比較のために左上に示した。図7は高年層、図8は若年層である。図7の高年層の結果を見ると、最も有声性によってデータが明確に分かれるのはvDelayを追加した場合(右上)

であることが視覚的に分かる。特にVOTが10～20msecあたりの区別が良くなっている様子が見える（予測確率の値が有声音で1.0に近いものが、他のモデルに比べて多い）。それに対し、iGapやpitchをパラメータとして追加したモデル（左下、右下）ではその辺りが改善されているようには見えない。逆にpitchを追加したモデルではVOT10msec付近の無声音で予測確立の低いものが増えているようにさえ見える。

一方、図8の若年層の結果を見ると、ピッチ以外の組み合わせのモデルでは、特にVOTが10msec以下のデータで、予測確率が実際の有声性とくいちがいがい、うまく判別できていないデータが見られる。これに対し、pitchを追加したモデル（右下）ではその部分の無声音は無声音、有声音は有声音（無声音でない）として予測されるであろうもの（予測確率がそれぞれ0.5より高いもの、あるいは低いもの）が多くなっているように見える。若年層は、先に3.2.1節で見たように、VOTとpitchの組み合わせのAICが最も小さかったが、この内容としてピッチが特にVOT値の小さい音声の判別に有効に働くことによるものであることが分かる。

次にこの2つのパラメータによるモデルの中で、それぞれの世代でもっとも当てはまりの良い



東北の話者における有声性に関わるパラメータの探索とその世代差 (高田)

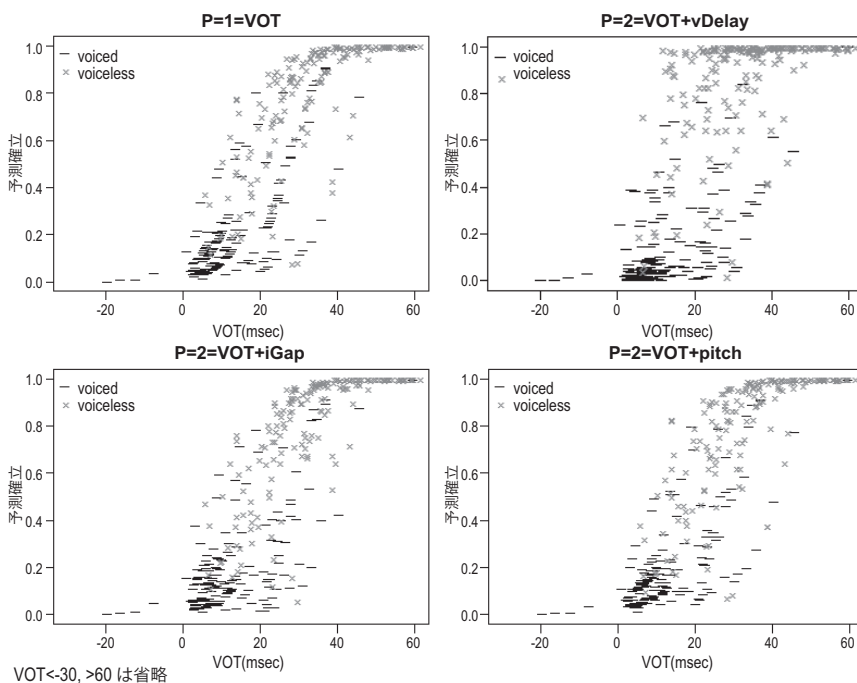


図7 高年層・2つの音響パラメータモデルの実データへの当てはめ (予測確立)

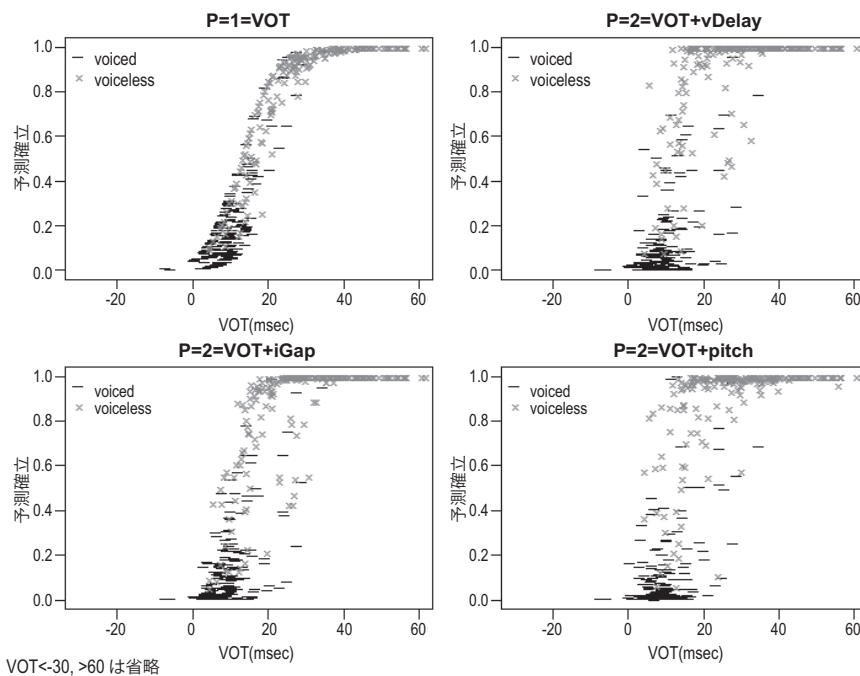


図8 若年層・2つの音響パラメータモデルの実データへの当てはめ (予測確立)

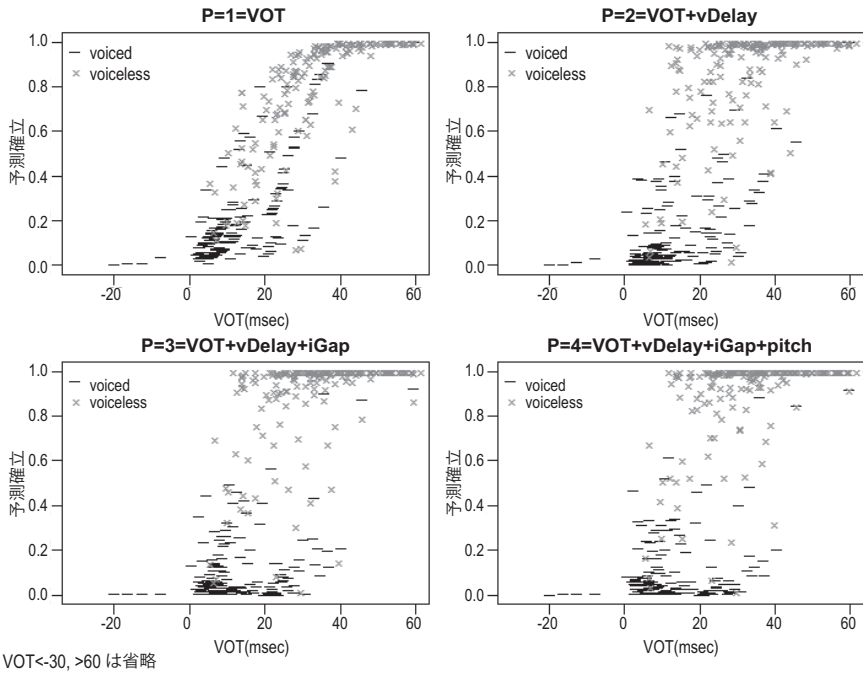


図9 高年層・音響パラメータモデルを順に増やしたモデルのデータへの当てはめ (予測確立)

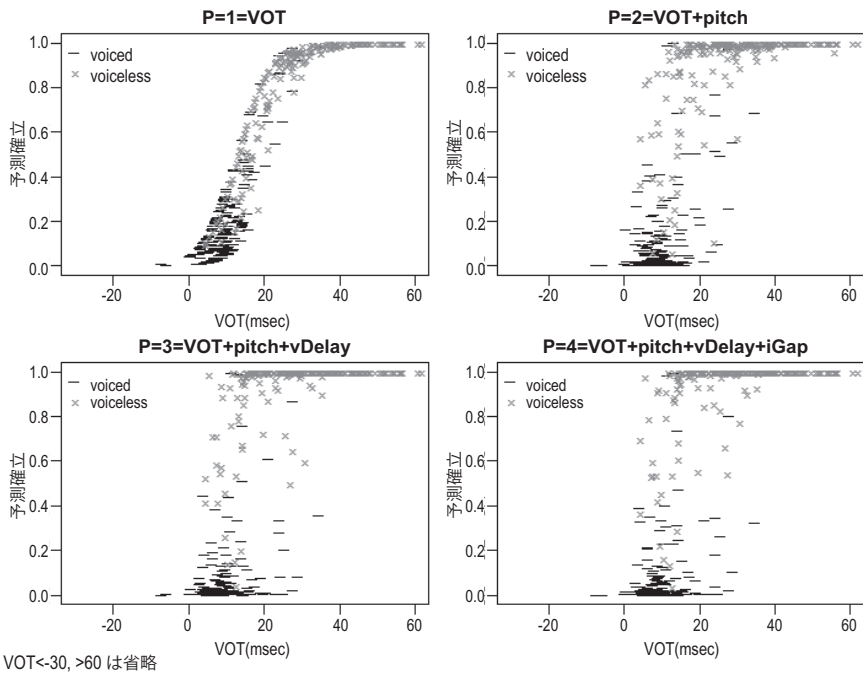


図10 若年層・音響パラメータモデルを順に増やしたモデルのデータへの当てはめ (予測確立)

モデルに、さらに残りのパラメータを加えていったモデルの適用を見てみよう。つまり高齢層の場合は VOT+vDelay のモデルを基本として iGap, そしてさらに pitch を加える。若年層の場合は VOT+pitch のモデルに vDelay, そしてさらに iGap を加える。

図9は高年層の結果である。これを見ると、VOT+vDelay にさらに iGap を加えたモデル（左下）では、区別の精度が全体に上がり、無声音でありながら予測確率の低かったものが上がると同時に、有声音でありながら予測確率の高かったものが下がったように見える。pitch を加えたモデルでは、その傾向が若干、さらに進められたとも見えるが、はっきりはしない。結果的に、パラメータが増えることで予測確率と実際の有声性の区別が良くなっていくと見ることができよう。

図10は若年層の結果である。3.2.1節で見たように、pitch 以外のパラメータは AIC をそれほど下げていなかった。図10をみると、細かく見れば、vDelay, iGap を追加することで有声音で VOT が 20~30msec 辺りのデータの予測が良くなっている（予測確率が高かったものの値が低くなっている）ように見える。しかし VOT に pitch を追加した場合ほど分布が大きく変わっておらず、貢献は部分的であると見える。

以上、各モデルのデータへの当てはめを通してパラメータの働き方を見たところ、高年層では各パラメータが順次、少しずつ全体の予測の精度を上げていくように見えた。一方、若年層では VOT と pitch でほぼ全体の予測確率が決まり、その他のパラメータは取り残された（予測確率が実際の有声性と食い違う）ある特定の条件のデータの判別を改善するといった部分的な働き方をしてるように見えた。どのパラメータが有声性の区別に関わるかというだけでなく、その関わり方についても、世代間での違いがあるようである。

#### 4. 考察

本稿では有声性に関与する音響的パラメータとして VOT, pitch, iGap, vDelay という4つのパラメータについて検証した。どちらの世代でも

VOT は有声性に大きく関与することが、単純な分布でもまたモデルからも確かめられた。しかしその他のパラメータとして何が大きく関与するかは世代間で異なっていた。若年層では VOT 以外には pitch が大きく関与することが、単純な数値の分布からも、またモデルの当てはまりの良さからも確認できるが、他のパラメータのモデルへの貢献は部分的であると見ることができよう。

一方高年層では VOT をメインとしながら、pitch よりもむしろ vDelay や iGap といった別のパラメータの関与が見られ、これら複数のパラメータが複合的な関与が見られた。またそれらのパラメータの値の分布は一見して分かるような単純なものではなく、調音位置との交互作用もあることから、より複雑に関与し合っていると考えた方がよい。

若年層で大きく関与することが確かめられた pitch が、高年層ではあまり関与していなかったということについては、Gao and Arai (2019) や邊 (2019) など先行研究で指摘されたピッチの関与が、若年層に限定される可能性を示している。邊 (2019) によれば、東北は特に有声性に pitch が大きく関与する地域であるという。東北は VOT に関して高年層において既に有声音がプラスの値を取る地域であり、このことから、ピッチの利用も高年層から見られる可能性があった。しかし本稿の分析の結果からそうではないことがわかった。つまり、東北以外の地域も含めて若年層に見られる有声性におけるピッチの関与は、「東北の伝統的な発音方法」が広がったものではない。若い世代で新たに生じた現象であると考えべきである。

#### 5. 今後の課題

以上のように東北の話者においては、世代によるパラメータの違いがあった。しかし他の地域で世代差があるのか否か、また各世代でのパラメータの関与の仕方については不明である。今後他の地域についても検証を進める必要がある。

なお両世代とも、本稿のモデルでは実際の有声性が正しく予測されない（予測確率が有声音で高いあるいは無声音で低い）データが残った。他に

考慮すべきパラメータがある可能性は大いにある。また今回検証に用いたパラメータについても解釈や妥当性の再検討が必要である。

今回検討しなかったが今後ぜひ検討に含めるべきこととして、まず挙げられるのはアクセント、音節構造である。本稿の分析語は1音節語と2音節語を含み、2音節語については1型アクセントであったが、この点を分析に含めることができなかった。また本稿で扱ったものは無意味語であったが、今後は当然ながら有意味語も分析すべきである。

再検討を要するパラメータとして、本研究で設定した  $vDelay$  については、改めて調音上の意味についての再解釈が必要である。そもそも、今回の検討では単純に  $F_1$  と  $F_2$  が出そろわないことをもって母音開始の遅れと考えたが、これはフォルマントの次数よりも、エネルギーの相対的な強弱とその周波数帯に意識を向けるべき現象だったのかもしれない。

なおこの  $vDelay$  が生じる理由について、現時点で筆者は3つの可能性を考えている。まず、そもそも喉頭音源のスペクトル傾斜は周波数が高いほどエネルギーが小さいために、出始めの弱いエネルギーで高次のフォルマントのエネルギーが検出されにくくなるという可能性が考えられる。つまり遅れが見られるのが高次フォルマントであるとすれば、その区間が長いというのは、それだけ出だし喉頭音源エネルギーが弱い音声だからだ、という解釈が可能かもしれない。

あるいはまた、低い周波数のエネルギーが相対的に強められているという可能性も考えられる。東川雅彦・竹中洋 (2003) によれば、ささやき声 (= 声門開放条件) において、有声音を発しようとする場合には無声音様の発声に比べて舌全体および喉頭を下げることで前庭腔を広げ、結果的に2000Hzより低い周波数のエネルギーが相対的に強くなるという。このような声道形状の変化が、ささやき声でなくても (通常の発話でも) あるとすれば、有声音では低周波数のエネルギーが相対的に強くなる可能性がある。声帯振動開始部で、特に2000Hzより下の帯域のエネルギーが見えて

も高周波域のエネルギーが現れないのは、相対的に高周波域のエネルギーが弱く高次フォルマントがスペクトログラム上の表示に至らない状態であるということも考えられる。ただし、本稿の測定で  $F_1 \cdot F_2$  開始の遅れは無声音の方が大きい傾向を見せ、有声音が無声音を上回る分布を見せたのは唯一高年層の軟口蓋音においてのみであった。こうした点で、これも単純に当てはめられるとは思われない。

あるいは、もしかすると喉頭の緊張や氣息性といった音声のレジスターに関する可能性もある。筆者はすでに一度、これに関係する  $H1-H2$  というパラメータを用いた分析を行っているが (Micko Takada, Eunjong Kong, Kiyoko Yoneyama and Mary Beckman 2017), その際はあまり明確な結果が得られなかった。しかし再度、計測ポイントや手法を再検討し、改めて分析を試みるべきかもしれない。以上のように、考えるべき点は多々あり、測定方法の改善も含め今後検証が必要である。

また  $iGap$  について、本研究では破裂から声帯振動開始までの平均値を計算に用いたが、これについても破裂のエネルギーの強さを表す指標とするためには、より短い区間を設定すべきだったかもしれない。

上記のように、本稿は使用するパラメータの検証という点でもまだ検討が必要である。しかし少なくとも、東北において、有声性に関わる音響パラメータに世代差が存在することは確認され、これは重要な指摘であると考えられる。今後検証を重ね、他地域、他世代を含め、現代日本語の有声性に関わるパラメータの全体像を明らかにしていきたい。

#### 謝辞

本稿は2019年11月30日に日本音声学会第340回研究例会で発表した内容に基づいている。発表の際に様々なご教示をいただいた。この場を借りて御礼申し上げる。

本稿で分析した資料は2006~2007年に多くの話者の皆様のご協力により収集したものである。当時録音調査にご協力いただいた話者の皆様に改めて感謝申し上げます。なおその際、財団法人博報児童教育振興会に

よる「2005年度 第1回博報『ことばと文化・教育』研究助成」を受けた。また現在研究を進めるにあたり JSPS 科研費 JP19K00634の助成を受けている。

#### 注

- 1) 本稿のデータでは /data, tada/ についてはモーラ数が2モーラで他は1モーラであるため検討に入れるべきであった。実際これをパラメータにした試みも行ったところ、特に若年層で効果が高かった(後ほど示す M3に加えた場合、AICは高年層: 207.97, 若年層: 95.85)。
- 2) 邊 (2019) の VOT とピッチ (fo) のどちらが一時的特徴かという議論で用いられているのは判別確率であり AIC ではない。本稿のモデルに関しても判別確率を求めて比較したが、若年層において、VOT 単独で91%、ピッチ単独で84%であり、やはり VOT の方が高い。

#### 引用文献

東川雅彦・竹中洋 (2003) 「さきやき声における /pa/ と /ba/ の出し分け: ヘリカル CT による声道形態の検討」『音声言語医学』44-4, 321-325.

邊姫京 (2019) 「日本語における語頭閉鎖音の音響特徴: VOT と後続母音の fo」『音声研究』23, 174-197.

Gao, Jiayin and Takayuki Arai (2019) “Plosive (de-)voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery”, *Journal of Phonetics* 77, 100932.

Shimizu, Katsumasa (1996) *A Cross-language Study of Voicing Contrasts of Stop Consonants in Asian Languages*, Seibido.

高田三枝子 (2011) 『日本語の語頭閉鎖音の研究: VOT の共時的分布と通時的変化』くろしお出版

Takada, Mieko, Eunjong Kong, Kiyoko Yoneyama and Mary Beckman “Individual differences in the phonetic realization of the Japanese stop-voicing contrast” The 4th International Workshop on Sound Change, University of Edinburgh, Edinburgh, U.K., 19-22 April, 2017.

Tamura, S., Kazuhito Ito, Nobuyuki Hirose and Shuji Mori (2018) “Psychophysical Boundary for Categorization of Voiced-Voiceless Stop Consonants in Native Japanese Speakers”, *Journal of Speech, Language, and Hearing Research* 61, 789-796.