

原 著

合成音声を用いた[s]の口蓋化構音に関する検討

荒井 隆行¹⁾ 岡崎 恵子²⁾ 今富 摂子³⁾

要 約 : [s]の口蓋化構音の音響的性質を裏付けるため、摩擦音を合成し聴取実験を行った。健常男児による正常構音の[stu]において、[s]の部分を実の各モデルによる合成音と置き換えて音声資料を作成した。第1のモデルは帯域通過フィルタであり、帯域値は一定、遮断周波数を段階的にシフトさせた。第2のモデルは、2次の全極型フィルタであり、極の周波数を段階的にシフトさせた。第3のモデルは、[s]の口蓋化構音に対する高次の全極型フィルタである。各フィルタは、白色性雑音で駆動した。9人の言語治療士に対して聴取実験を行い、結果を[s], [ʃ], [s]の口蓋化構音、その他の4項目に整理した。その結果、第1のモデルは[s]の口蓋化構音の合成に適していないこと、第2のモデルで合成された摩擦音のうち、極の周波数が2~3 kHzにある場合に[s]の口蓋化構音として聴取される傾向が強いこと、第3のモデルにおいて2 kHz付近に存在する極の帯域幅が500 Hz程度、もしくはそれ以下の場合に[s]の口蓋化構音として聴取される傾向が強いことが分かった。

索引用語：口蓋化構音, 音響分析, 合成音声, 摩擦音, 口蓋裂

Palatalized Articulation of [s] Sounds Using Synthetic Speech

Takayuki Arai¹⁾, Keiko Okazaki²⁾, Setuko Imatomi³⁾

Abstract : Using several models for synthesizing speech, we tested human perception of certain synthetic sounds to verify already published characteristics of palatalized articulation (PA). In the mono-syllable [stu] uttered with normal articulation, part of the fricative sound [s] was replaced by synthetic noise. The following three models were used to synthesize [s]. The first model is a bandpass filter which has a pass band in a specific frequency range. The cutoff frequencies were shifted in intervals from low to high, while the bandwidth was fixed. The second model is an all-pole model with second-order linear predictive(LP) analysis. To implement this filter partial correlation(PARCOR) coefficients were used. The frequency of the Poles was shifted in intervals from low to high. The third model is an all-pole model using higher order LP analysis for a typical PA of [s]. Each

¹⁾ 上智大学理工学部電気電子工学科：〒102 東京都千代田区紀尾井町7-1

²⁾ 昭和大学医学部形成外科：〒142 東京都品川区旗の台1-5-8

³⁾ 昭和大学藤が丘リハビリテーション病院言語室：〒227 神奈川県横浜市青葉区藤が丘2-1-1

¹⁾ Department of Electrical and Electronics Eng., Sophia University : 7-1, Kioi-cho, Chiyoda-ku, Tokyo 102

²⁾ Department of Plastic and Reconstructive Surgery, School of Medicine, Showa University : 1-5-8 Hatanodai, Shinagawa-ku, Tokyo 142

³⁾ Department of Speech Therapy, Fujigaoka Rehabilitation Hospital, Showa University : 2-1-1, Fujigaoka, Aoba-ku, Yokohama-shi, Kanagawa 227

原稿受理：1994年10月20日

filter is excited by white noise to synthesize the fricatives. The hearing discrimination of nine speech therapists formed the data for the perceptual experiment ; they were each requested to indicate what they heard. Their replies were categorized as : “[s]”, “[ʃ]”, “the PA of [s]”, or “other”. From the results we concluded : 1) the first model is not appropriate for synthesizing the PA of [s] ; 2) fricatives, which have a peak in the range of 2-3kHz, tend to be identified as the PA of [s] when synthesized by the second model : and 3) fricatives synthesized by the third model using sixth or higher order LP tend to be identified as the PA of [s].

Key words : palatalized articulation, acoustic analysis, synthetic speech, fricative sound, cleft palate

はじめに

口蓋化構音は、歯音、歯茎音の構音点が口蓋に後退することによって舌背の中央部と口蓋でつくられる歪み音である¹⁾。口蓋化構音の動態は、レントゲン映画、ダイナミックパラトグラフの使用、発話時の構音器官の観察によって解明されている²⁾。

口蓋裂言語の音響分析は母音のフォルマント分析を行ったもの³⁻⁵⁾、子音に関するもの⁶⁻⁷⁾が報告されているのに対し、[s]の口蓋化構音に関するものは少ない⁸⁻¹¹⁾。[s]の口蓋化構音の音響特性は、サウンドスペクトログラフによるセクション分析から、[s]の口蓋化構音のエネルギーの集中は4 kHz以下に観測され、[s]や[ʃ]の正常構音とはことなっていることが確認されている⁹⁾。しかし、それらの報告で確認された音響的性質が口蓋化構音に特徴的であるかどうかを聴取実験によって確認した報告はまだない。

本論文では、その音響分析の結果を裏付けるため摩擦音を合成し、エネルギーの集中は4 kHz以下であるという音響的性質が口蓋化構音に特徴的であるかどうかを聴取実験によって確認した。

摩擦音の合成

先行研究および予備実験によって、エネルギーの集中する周波数帯域が正常構音の[s]は5 kHz以上、[s]の口蓋化構音は1~4 kHz、正常構音の[ʃ]は3~7 kHzと明らかにされていることから⁹⁾、それらをカバーする周波数範囲において摩擦音を合成した。摩

擦音の合成方法には、以下の3つのモデルを用いた。いずれについても、摩擦音は各フィルタを白色性雑音で駆動することによって合成した。

1. 帯域通過フィルタによるモデル

第1のモデルとして、ある特定の周波数帯域を通過帯域とする帯域通過フィルタを用いた。フィルタの帯域幅は3 kHzに固定しておき、遮断周波数を低周波域から高周波域まで段階的にシフトさせた(表1)。フィルタは、Hamming窓を用いて256タップのFIR (finite impulse response)フィルタで実現した。

2. 2次の全極型フィルタによるモデル

第2のモデルは、線形予測法による全極型フィルタを用いる方法である。2次のモデルを考え、極の周波数を低周波域から高周波域まで段階的にシフトさせ、フィルタを設計した(表2)。極の帯域幅は100 Hzに固定し、フィルタは偏自己相関係数によって実現した。

3. 高次の全極型フィルタによるモデル

第3のモデルでは、典型的な[s]の口蓋化構音に対して線形予測分析から得られた偏自己相関係数による全極型フィルタを用いた。その際、分析次数は2から8次まで段階的に変化させた(表3)。

実験方法

音声資料は、健常男児(6歳)が発話した正常構音の単音節[su]の摩擦子音[s]の部分合成音声の[s]と置き換えて作成した。聴取実験の信頼性を高めるため、その健常男児が発話した[s]および口蓋裂児(5歳、左唇顎口蓋裂)が発話した[s]の口蓋化構音も資料に含

表1 帯域通過フィルタによるモデルA (添え字は通過帯域 [kHz])

モデルA	A ₁₋₄	A ₂₋₅	A ₃₋₆	A ₄₋₇	A ₅₋₈
下限の遮断周波数 [Hz]	1000	2000	3000	4000	5000
上限の遮断周波数 [Hz]	4000	5000	6000	7000	8000

表2 2次的全極型フィルタによるモデルB (添え字は極の中心周波数 [kHz])

モデルB	B _{2.0}	B _{2.5}	B _{3.0}	B _{3.5}	B _{4.0}	B _{4.5}	B _{5.0}	B _{5.5}	B _{6.0}
極の中心周波数 [Hz]	2000	2500	3000	3500	4000	4500	5000	5500	6000

表3 高次的全極型フィルタによるモデルC (添え字は分析次数)

モデルC	C ₂	C ₄	C ₆	C ₈
分析次数	2	4	6	8

め, 2節で述べた18種類の合成音声を合わせて合計20種類の単音節[sw]を作成した。[s]は[s]の口蓋音といわれていることから, [ʃ]も実験の対象とした。音声の分析, 編集および合成は表4に示す条件下においてSUN Sparc Station上でを行い, 聴取実験用の音声資料は最終的にカセットテープに録音した。

聴取実験は9人の言語治療士に対してを行い, 20種類の単音節を各3回ずつランダム提示した。実験は刺激提示順序を変え2回行った。回答方法は, 録音された単音節を聞いた後に摩擦音に対して表5に示す選択肢のいずれかを選択するものとした。

結 果

言語治療士9人の聴取実験の回答のうち, [s]と「[s]に近い」, 「[ʃ]に近い」, 「[s]の口蓋化構音」と「[s]の口蓋化構音に近い」というそれぞれの組は, どちらかを選択する際に被験者によって選択の基準にばらつきがあるものと考えられたので, その影響を減らすため, 「[s]に近い」を[s]に, 「[ʃ]に近い」を[ʃ]に, 「[s]の口蓋化構音に近い」を「[s]の口蓋化構音」にそれぞれを統合して集計を行った。結果のうち, 正常構音の[s]を「正常構音の[ʃ]」または「[s]の口蓋化構音」と回答したものは, [s]の口蓋化構音を「正常構音の[s]」または[ʃ]と回答したものはなかった。各モデルに対する結果を, それぞれ図1~3に示す。

まず図1から, A₁₋₄はほとんど「その他」に判定されていることが分かった。一方, A₂₋₅およびA₃₋₆では正常構音の[ʃ], A₄₋₇およびA₅₋₈では[s]に判定されることが多くあった。つまり, エネルギーの集中が2~6 kHzにある場合は[ʃ]に, 4~8 kHzにある場合は[s]に聴取される傾向にあることがわかった。このことは, ほぼ先行研究の結果と一致している。しかし, モデルAにおいて[s]の口蓋化構音として判定されることはほとんどなかった。

次に図2から, B_{2.0}およびB_{4.0}が[s]の口蓋化構音,

表4 実験条件

標準化周波数	16 kHz
量子化ビット数	16 bit
合成摩擦音の duration	200 ms

表5 聴取実験における回答の選択肢

選択肢
[s]
[s] に近い
[ʃ]
[ʃ] に近い
[s] の口蓋化構音
[s] の口蓋化構音に近い
その他

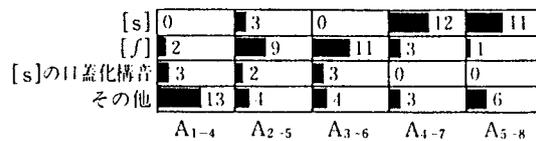


図1 モデルAに対する聴覚判定のヒストグラム

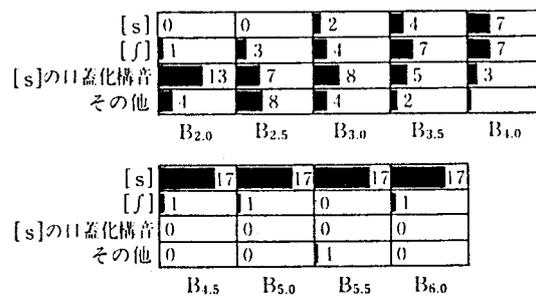


図2 モデルBに対する聴覚判定のヒストグラム

B_{3.5}が[ʃ], B_{4.5}, B_{5.0}, B_{5.5}およびB_{6.0}が[s]と判定されることが多いことがわかった。B_{2.5}の聴覚的判定では, 「その他」と[s]の口蓋化構音に多く分類され, [ʃ]にも少し分類された。また B_{4.0}に関しては, [s]と[ʃ]に同数に分類されている。つまり, 2次的全極型フィルタで摩擦音を合成した場合に極の周波数が2.0~3.0 kHzにある場合は[s]の口蓋化構音に, 3.5~4.0 kHzにある場合は[ʃ]に, 4.0 kHz以上にある場合は[s]に聴取さ

[s]	6	7	2	0
[ʃ]	0	0	0	1
[s]の口蓋化構音	3	1	13	16
その他	9	10	3	1
	C ₂	C ₄	C ₆	C ₈

図3 モデルCに対する聴覚判定のヒストグラム

れる傾向にあることがわかった。これも、先行研究の結果とよく一致している。

最後に図3から、C₆およびC₈が[s]の口蓋化構音と判定されることが多いことがわかった。

考 察

以上の聴取実験から、[s]の口蓋化構音を合成するにはモデルAのように帯域の広いノイズは適しておらず、モデルBのように帯域幅の小さいノイズの方が適していることがわかった。摩擦音に関しては多くの研究があるが¹²⁾、Manriqueら¹³⁾は帯域通過フィルタを用いてエネルギーの集中域の違う雑音を作成し、摩擦音の識別を行ったところ、5～8 kHzにエネルギーのピークがある場合に[s]、2.5 kHz付近にピークがある場合に[ʃ]に識別されたという報告をしている。その際、帯域幅は2 kHzのものを用いている。モデルAにおいてはフィルタの帯域幅を3 kHzとしたが、これは予備実験によって2 kHzを用いた場合には[ʃ]に判定されなかったこと、また4 kHz以上を用いた場合にはスペクトルの白色性が強まり、さらに[s]や非言語的雑音としての聴覚印象を与える可能性が強かったことによる。モデルBについても、帯域幅を広くすればそのスペクトルの白色性が強まるため、帯域幅はあまり広くすることは望ましくなかった。

鼻咽腔閉鎖機能が良好な場合に発話された母音は全極型モデルで近似されることが知られており、多くの場合は線形予測分析が一般的に用いられている。一方、鼻咽腔閉鎖機能が良好であっても音源が口腔の前方に存在する摩擦音などは、一般に零点を有するために全極型モデルでは完全には表現できない。しかし、分析次数を高くすれば全極型モデルで近似することは可能である。モデルCの実験において、[s]の口蓋化構音を高次の全極型フィルタで近似した場合、次数を6次もしくは8次以上、すなわち周波数領域で極の数を3もしくは4以上にしたとき[s]の口蓋化構音に聴取される傾向にあった。このことを詳細に考察するため、モデルCによる合成摩擦音に対し1～4 kHzの範囲に存在する極の中心周波数と帯域幅を求めた。おのおのの次数に対して求めた結果を、表6に示す。この表をみると、次数が上がるにしたがって帯域幅が小さくなって

表6 モデルCによる合成摩擦音において1～4 kHzの範囲に存在する1極の中心周波数と帯域幅

モデル	中心周波数 [Hz]	帯域幅 [Hz]
C ₂	2248	2158
C ₄	1749	1024
C ₆	2334	511
C ₈	2160	263

いるのがわかる。また、次数が6次以上の場合に帯域幅が約500 Hz以下になっている。これらのことから、2 kHz付近に存在する極の帯域幅が500 Hz程度、もしくはそれ以下の場合に[s]の口蓋化構音として聴取される傾向が強いことがわかった。

Heinzら¹⁴⁾によれば1つの極と1つの零点だけによるシステムを白色性雑音で駆動することによって、摩擦音に対する良い近似を得ている。そして、極の周波数が3 kHz以下であるときに[ʃ]、4～8 kHzであるときに[s]に識別されたという報告をしている。また、Flanagan¹⁵⁾は2つの音響管の接続によって摩擦子音をモデル化している。また白井ら¹⁶⁾は、計算機シミュレーションによって肺、喉頭、声道などを制御し、乱流雑音を発生させて摩擦子音を合成している。本論文で用いたモデルは非常に単純なものであるが、これをさらに改良するためには乱流雑音の発生メカニズムを根本的に考える必要がある。

本実験で用いた単音節[su]において、母音[u]の部分は健常児によるものである。その健常音声である後続母音の影響で摩擦音が口蓋化構音として聴取されにくい可能性があった。Harris¹⁷⁾はアメリカ英語のCV音節に対して、先行する子音が[s]と[ʃ]の場合には両者の相違は子音部にあり、[f]と[θ]の場合には相違は有声部にあるとしている。本実験で用いた音声資料においては、フォルマント周波数の連続的な遷移など調音結合に関する考慮に欠けていることから、今後は合成方法を検討するうえでの課題とするとともに他の後続母音についても実験を行う必要がある。

おわりに

[s]の口蓋化構音の音響分析において確認された音響的性質が口蓋化構音に特徴的であるかどうかを裏付けるため、摩擦音を合成し聴取実験を行った。音声資料は、健常男児が発話した正常構音の単音節[su]において、摩擦子音[s]の部分を合成音声の[s]と置き換えて作成した。合成には、3種類のモデルを用いた。第1の

モデルは、ある特定の周波数帯域を通過帯域とする帯域通過フィルタである。フィルタの帯域幅は一定にしておき、遮断周波数を低周波域から高周波域まで段階的にシフトさせた。第2のモデルは、線形予測法における全極型フィルタである。2次のモデルを考え、極の周波数を低周波域から高周波域まで段階的にシフトさせ、偏自己相関係数によるフィルタを構成した。第3のモデルは、典型的な[s]の口蓋化構音に対して線形予測分析から得られた高次の全極型フィルタである。それぞれのフィルタに対し、白色性雑音で駆動することによって摩擦音を合成した。9人の言語治療士による聴取実験の結果、第1のモデルは[s]の口蓋化構音を合成するには適していないこと、第2のモデルで合成された摩擦音のうち、極の周波数が2~3 kHzにある場合に[s]の口蓋化構音として聴取される傾向が強いこと、第3のモデルにおいて2 kHz付近に存在する極の帯域幅が500 Hz程度、もしくはそれ以下の場合に[s]の口蓋化構音として聴取される傾向が強いことがわかった。

臨床において言語治療士によって行われている聴覚判定および構音器官の直接的観察に加え、サウンドスペクトログラムなどの一般的な音響分析装置を用いた口蓋化構音の分析および客観的な評価が実施されれば、診断が容易になると同時にその信頼性もより高いものになる。本研究によって、その指標となる音響分析を裏付けるのに十分な結果が得られた。

本研究の聴取実験にご協力頂いた言語治療士の先生方、ならびにご指導頂いた上智大学理工学部電気電子工学科吉田裕一教授に深謝いたします。

本稿は、第39回日本音声言語医学会総会(金沢)において発表した。

文 献

- 1) 岡崎恵子：口蓋裂言語の1型としての口蓋化構音。日形会誌, 2 : 164-176, 1982.
- 2) 岡崎恵子, 鬼塚卓弥, 阿部雅子, 他：口蓋裂における異常構音としての口蓋化構音について—ダイナミック・パラトグラフおよびX線映画による観察—。音声言語医学, 21 : 100-120, 1980.
- 3) Dickson, D. R. : An acoustic study of nasality. J. Speech and Hearing Research, 5 : 103-111, 1962.
- 4) Schwartz, M. F. : Acoustic measures of nasalization and nasality, Communicative Disorders related Cleft Lip and Palate edited by K. R. Bzoch, 194-200, Little Brown and Company, Boston, 1979.
- 5) 平野 実：口蓋裂音声の音声学的研究。耳鼻臨床, 56 : 22-66, 1963.
- 6) 今井 章：Sonagraphによる異常音声の研究第二編 口蓋裂形成術後の異常音声について。阪大歯誌, 4 : 578-600, 1959.
- 7) 吉田 稔：正常音声ならびに口蓋裂異常音声の周波数スペクトルに関する研究。口外誌, 6 : 527-544, 1973.
- 8) 岡崎恵子, 加藤正子：口蓋化構音の音響分析, [s]について。音声言語医学, 29 : 225-231, 1988
- 9) 和久本雅彦：異常構音の客観的評価法の開発—口蓋化構音の音響特性の検索—。日口蓋誌, 14 : 21-43, 1989.
- 10) 福田登美子：口蓋裂術後患者における口蓋化構音による異常音声の音響学的特徴と構音治療に伴う変化に関する研究。阪大歯誌, 38 : 8-35, 1993.
- 11) 小澤由嗣, 岡崎恵子：口蓋化構音の構音動態と音響的特徴の関連—症例間の差異の検討—。音声言語医学, 35 : 322-330, 1994.
- 12) Hughes, G. W., and Halle, M. : Spectral properties of fricative consonants. The Journal of the Acoustical Society of America, 28 : 303-310, 1956.
- 13) Manrique, A. M. B., and Massone, M. I. : Acoustic analysis and perception of Spanish fricative consonants. The Journal of the Acoustical Society of America, 69 : 1145-1153, 1981.
- 14) Heinz, J. M., and Stevens, K. N. : On the properties of voiceless fricative consonants. The Journal of the Acoustical Society of America, 33 : 589-596, 1961.
- 15) 白井克彦, 藤沢浩道, 小出雄悟：摩擦子音発声過程のモデル化。電子通信学会論文誌, 58-A No. 6 : 345-352, 1975.
- 16) Flanagan, J. L. : Speech Analysis ; Synthesis and Perception. Springer-Verlag, Berlin, Germany, 1972.
- 17) Harris, K. S. : Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech, 1 : 1-7, 1958.

別刷請求先：〒157 東京都世田谷区成城2-5-18

荒井隆行