

TR-I-0354

逐次状態分割法 (SSS) による日本語
子音に関する一検討

A Study of Japanese Consonants through the Successive

State Splitting (SSS) Algorithm

鷹見淳一

嵯峨山茂樹

Jun-ichi TAKAMI

Shigeki SAGAYAMA

衣笠和江†

壇辻正剛†

Kazue KINUGASA

Masatake DANTSUJI

1993.3

概要

本報告では音声認識の手法として開発された逐次状態分割法 (SSS - Algorithm) を日本語の子音の分析に応用して、その結果得られた隠れマルコフ網の構造を音声学的な観点から検討を試みた。その成果として、子音の調音位置が前後の音韻の特性によって微妙に変動したり、語頭と語中で調音機構の特性が変化するなどといった従来の調音音声学的な知見に対して、音響音声学的な側面からも確認することが可能になった。

†関西大学

Kansai University

ATR 自動翻訳電話研究所

ATR Interpreting Telephony Research Laboratories

© (株)ATR 自動翻訳電話研究所 1993

© 1993 by ATR Interpreting Telephony Research Laboratories

目次

1	はじめに	2
2	音素の概念	2
3	SSS による子音の分割状態	6
4	おわりに	10

1 はじめに

本研究では、音声の自動認識の研究の過程で開発された“Successive State Splitting(SSS) Algorithm”-「逐次状態分割法」の研究成果を取り入れながら、調音音声学で得られた知見に基づいて、音素の異音の生起や共調音(co-articulation)の問題を中心に検討を加えてみたいと思う。また、弁別素性の観点からも音響的特性がどのように反映されているのかということについても検討を試みていきたいと思う。

音素は、具体的な形をとって明らかに現れる時に環境によって大きく影響を受けて変形する。このような変形は、当該音素、先行音素、後続音素、アクセントの有無、語頭からの位置などの影響による「位置的変異音」としての「異音(allophone)」、話者、発話速度、ピッチ、パワー、感情などにも起因する「自由変異(free variant)」や「変音(variphone)」、言語間や方言間の相違を扱う「類音(diaphone)」などの言語学的用語にも反映されている。本研究は、これらの現象に関して、音声認識の成果を利用してアプローチを試みたものである。

現在の音声認識技術では隠れマルコフモデル(Hidden Markov Model:HMM)を用いた音声認識が有望視されている。しかし、HMMを用いて高精度で頑健な音素モデルを形成する場合、以下の様な問題点が指摘されている。

- 1) 音素環境の分類をどのように行えばよいか？
- 2) 各 HMM の表現能力をどの程度に設定すればよいか？
- 3) HMM の増加に伴う 1 モデル当たりの学習サンプルの減少をどのように克服すればよいか？

これらの問題点を解決し、高精度で頑健な HMM を生成するための手段として考案された手法が「逐次状態分割法(Successive State Splitting:SSS)」である。この SSS によって実現される表現効率の高い音素環境依存 HMM が隠れマルコフ網(Hidden Markov Network: HMnet)である。

本研究では日本語の子音音素を取り上げながら逐次状態分割法(SSS)によって自動的に生成された HMnet の構造を利用して音声学的な知見へのアプローチを試みたいと思う。

2 音素の概念

ここで、音素の概念について簡単に解説しておこう。意味の区別に関係する音が音韻であるが、アメリカ構造言語学では「音声的に類似した音」で「意味を区別する最小の単位」を「音素」と呼び、//で囲んで示すことになっている。例えば、日本語の撥音を例にとって考えて見よう。

サンバ「産婆」
サンダイ「三台」
サンカ「参加」

上の三語を見ると「ン」はすべて同じ音に思われるかもしれない。しかし、各単語を発音して注意深く観察してみるとその相違に気付くであろう。「サンバ」では「ン」の時点ですでに唇が閉じられていることに気付くであろう。もちろん唇を閉じずに発音することも可能である

が、普通に発音すると必ず唇が閉じられてしまうであろう。一方、「サンダイ (三台)」や「サンカ (参加)」は「ン」の時点では唇が閉じられていないことに気付くであろう。「サンダイ」と「サンカ」の「ン」をさらに注意深く観察すると「サンダイ」の「ン」では舌先や舌端が歯もしくは歯茎に接しているのに対して、「サンカ」の「ン」では舌の先が下がって後(奥)舌がもり上がり上顎の奥の方(軟口蓋)に接しているのことに気付くであろう。これらを音声記号で表記すると次のようになる。

サンバ [samba]
サンダイ [sandai]
サンカ [saŋka]

さらに注意してみると、一つの単語の内部でも変化しているのことに気付くであろう。例えば、「本」という単語を例にとってみよう。

「本」
「本も買いました」の「本も」
「本とノートを買いました」の「本と」
「本かノートを買いなさい」の「本か」

という部分を声に出して発音してみると、「本も」、「本と」、「本か」のそれぞれの「ン」の部分の発音は各々「産婆」、「三台」、「参加」の「ン」の発音と並行しているのことに気付くであろう。音声記号で表すと以下のようになる。

本も [hommo]
本と [honto]
本か [hoŋka]

一方、単独に発音した「本」の「ン」では舌の接触到に気付かないかもしれない。この音は [N] で表される場合が多い。さらにこれらことから、撥音の発音の相違は後続音素に影響されたものであるのことに気付くであろう。日常で意識されることもなく無意識の内に発音し分けているのであり、日本語の話者には同じ単位、すなわち「撥音」としてとらえられているのである。そして、仮に、唇を閉じずに「サンバ」の「ン」を発音してもそれが他の意味のコトバになるわけではなく、このような音声的な差違意味の区別に役立っていない。すなわち「弁別的」でないのことに気付くであろう。そこで、これらの撥音は一つの音素として取り扱われることになり、斜線で囲んで /N/ で表すことになるのである。そして、撥音 /N/ がそれぞれの環境で実際に現れる形の [m],[n],[ŋ] などは異音 (allophone) と呼ばれるのである。特にこの場合は後続音素により [m] が現れるか [n] や [ŋ] が現れるかが規定されるので条件異音と呼ばれる。また、この場合は一定の相補い合う条件下でのみ現れるので「相補分布をなす」という。さらに、この場合は後続音素と同じ調音位置を持つようになるが、このように前後の音と同じ特徴を共有するようになる現象を「同化」(assimilation)と呼んでおり、この場合のように先行する音が後続する音に影響されるような場合を特に「逆行同化」(regressive assimilation)と呼んでいる。

ただし、「音素」の定義に関しては各学派で少しずつ異なっている。音素 (phoneme) という言葉はフランスの言語学者でパリ言語学会の創立委員の一人でもある Dufrique-Desgenettes によって 1873 年にドイツ語の Sprachlaut の訳語として作られたとされている。一方、Daniel Jones(1957) は 1876 年に Louis Havet によって創られたとしているが、一般には 1873 年の Dufrique-

Desgenettes 説が有力である。いずれにしても、当初は、今日でいう言語音、あるいは文節音という意味合いで使われていたが、この用語が構造言語学の祖と称される Saussure や Courtenay, あるいはプラーグ学派の人々に採用され、次第に現在使われているような性格を与えられていくのである。

Baudouin de Cortenary は当初語源的に関係をもつ形態の交替の基底にある単位として音素の概念を用いていた。その後、音素を心理的単位と考え、音声の世界に属する同一観念であって、同じ音が異なって発せられることによって生じる印象が心理的に融合されて心の中に生成されるものとして定義するようになった。Saussure は音素は対立的、相対的かつ否定的な実体であるとしたが、まだ音素を弁別的な機能に関する音声単位に限定することなく用いたりもしていた。プラーグ学派も当初音素を心理的な単位としていたが、機能的な定義へと変化していき、当該の言語においてより小さい連続的音韻的単位に分析できない音韻的単位として定義するようになった。また、音韻的に関与的な特徴の総和であるとも規定している。ある言語において、知的意味を区別する働きをもつ音声的差異として音韻的対立 (phonological opposition) の概念を導入するようになり、Trubetzkoy は以下の様な対立の体系を設定した。

1. 体系内の他の成員に対する関係

- (a) 両辺的対立 (bilateral opposition) と多辺的対立 (multilateral opposition) : 両辺的対立は対立項が2つだけのものをいう。多辺的対立とは /p/, /t/, /k/ にみられるような調音点の対立の様に対立項が3つ以上のものをいう。
- (b) 孤立的対立 (isolated opposition) と並行的対立 (proportional opposition) : /r/ と /l/ の対立の様なある対立項の関係が他の対立項に見出せないような対立を孤立的対立という。

2. 対立する項の間関係

- (a) 欠如的対立 (privative opposition) : 対立項の一方が持っている特徴を他方が持っていないような対立。
- (b) 漸次的対立 (gradual opposition) : 同じ特徴の程度の差によって区別される対立。
- (c) 両極的対立 (equipollent opposition) : 特徴の有無や程度の差によっても区別されず、論理的に対等であるとみなされるような対立。

さらに、中和や原音素の概念等も導入された。一方、アメリカ構造言語学の祖と呼ばれるサビアは心理的型、交替および基底形を力説したが、もう一人のアメリカ構造言語学の祖と称される Bloomfield は音素を音声特徴の最小単位であり、意味を区別する最小の単位であると定義していた。Bloomfield は又、音素を「弁別素性の束」としても定義しており、純粋に物理的な特徴として音素を性格付けようとしていたが、後期ブルームフィールド学派では音を構成する特徴よりも音の類 (class) が重視されるようになった。

音素分析においては「相補分布」と「音声的類似性」の原則が重要視された。相補分布の概念は以下の様に説明される。たとえば、音節構造が CV 型で5母音しか持たないある言語において [s] と [ʃ] の分布が、[s] は母音 /a/, /e/, /o/, /u/ の前に現れ、[ʃ] は母音 /i/ の前にだけ現れると仮定してみよう。そうすると、[s] と [ʃ] の現れる環境は全く重なることなく、[s] の現れる環境では [ʃ] は現れないし、逆に、[ʃ] の現れる環境では [s] は現れないことになる。そ

のような分布を相補分布と名付け、その場合、[s]と[j]は一つの音素とみなされる。そして、音素として、/s/を設定し、[s]と[j]は音素/s/の「異音」と名付けられ、以下の様に定式化される。

$$\begin{array}{l} /s/ \rightarrow [j] \ / _ /i/ \\ \quad \quad [s] \ / \text{その他の場合} \end{array}$$

しかし、相補分布の概念だけでは色々と問題が生じてくる。例えば、英語の[h]と[ŋ]は相補分布をなす。[h]は語頭の位置に来るか、もしくは、強勢のない母音の後に来る。一方、[ŋ]は語末に来るか、あるいは、軟口蓋音/k/,/g/の前に来るか、強勢のある母音の後に来る。従って、両者の分布は完全に相補的である。そこで、両者を一つの音素の異音と考えて音素/h/を立てると以下の様な定式化が可能である。

$$\begin{array}{l} /h/ \rightarrow [h] \ / \# _ , \quad \check{V} _ V \\ \quad \quad [ŋ] \ / \# _ \left\{ \begin{array}{l} k \\ g \end{array} \right\} , \quad \check{V} _ V \end{array}$$

しかし、一般に言語学者はこの解釈を否定する。その理由は[h]と[ŋ]の間に「音声的類似性」がないということである。[h]は無声声門摩擦音であり、[ŋ]は(有声)軟口蓋鼻音である。音声的類似とは調音的特徴を共有するという意味に解釈されるが、この場合、両者に共通した音声的特徴は、肺流出機構の子音であるという点だけであり、調音的特徴は共有していないことになり、一つの音素とみなすわけにはいかないことになる。一方、先程の[s]と[j]の例では共に歯擦音(sibilants)と呼ばれる無声摩擦音であり、音声的類似の条件を満たしていることになり、一つの音素とみなしてもよいことになる。

Jonesは音声学者の立場から音素論を展開し、既にSweetの「簡易ローミック(Broad Romic)」と「精密ローミック(Narrow Romic)」の区別に音素と異音の区別への対応がある程度認められるとしており、「音素は、性質において関連があり、一単語内で、いかなる成員も他の成員と同じ音声環境に現れることが無いように用いられる、ある言語における音声の一族である。」と定義した。Jonesにとって重要であったのは言語の表記、語学教育への音素論の応用であり、理論よりはむしろ実用面であった。しかし、Twaddellは心理的実在としての音素やBloomfield、Jones等の物理的定義を批判し、音素は抽象的な仮構の単位であるとした。Twaddellは語頭の/p/や語末の/-p/を別々に取り出して、これを「小音素(microphoneme)」と名付け、小音素の総和を「大音素(macrophoneme)」と名付けた。これに対して、ソビエトの言語学者Saumjanは音素分析に観察の段階と構成の段階の二重の抽出段階を設け、音素は構成段階に属する仮説的要素であり、音声は音素が実質として表れたもので、さらに音声が対立によって区別された「音素質(phonemoid)」が音素の基質となるという二段階説を展開した。さらに具体音素と抽象音素等の概念を導入するが、異音は具体音素質に、音素は抽象音素質に対応すると考えられている。

また、ブラーグ学派では、音節もしくはモーラを単位として担う特徴を韻律的特徴として扱ったが、これに対応するのがBloomfieldの二次音素やブルームフィールド学派の超分節音素の概念である。超分節音素には接続(juncture)、強勢(stress)、声調(tone)、音調(intonation)、長さ(duration)などが含まれる。

このように、構造言語学では音素が音韻論の中心的単位であったので、音素論(phonemics)、あるいは音素理論(phoneme theory)という名称がよく使われたが、次第に音素が重要性を失ってきたので妥当性に欠けるようになり、最近では音韻論と言う名称の方が好まれるようになった。

た。生成音韻論では音素はもはや言語学的に有為な単位とは認められず、素性表示だけが言語学的に有意義であり、分節素 (segments) は慣習的な省略法であり、それ自身は言語学的な意義はないと結論づけられている。

日本語の音節構造において、一般に、子音が音節の頭位 (onset) に立ち、音節核を担う母音との結合が非常に密接であることと関連している。/k/ の調音点に注目すると、これは /k/ が後続する音素である母音や半母音と同化現象を起こして、調音点が少しづつ変化し、それに対応して音響上の差が生じてクラスタリングが行われたものと解釈される。すなわち、[+back, -low] の素性を持つセグメントの前で調音点が後寄り (口蓋垂寄り) の異音 [k] になり、[+front] の素性を持つセグメントの前では調音点が前寄りになって (硬) 口蓋化した [k'] になる。そして、その他の場合は軟口蓋音 [k] になるという現象を反映しているものと解釈することが可能である。

3 SSS による子音の分割状態

音声の自動認識の研究では、音素ごとに音響パターンを隠れマルコフモデル (Hidden Markov Model: HMM) とよばれる確率モデルでモデル化して認識を行なう方法が主流となっている。さらに、高精度の認識を実現するための手段として、音素コンテキストの影響を考慮した音素環境依存 HMM を用いる手法が有望視されている。一般には、あらゆる音素環境について別々のモデルを用意することは困難であるので、音素環境はいくつかのカテゴリに分類して扱う。しかし、従来は音素環境の分類をどのように行なうかというのが問題であった。

逐次状態分割法 (Successive State Splitting: SSS)[鷹見 1991, 鷹見 1992] はこの問題を解決するための手法として考案された方法である。SSS は、状態共有の考え方を取り入れた効率的な音素環境依存 HMM の表現法である隠れマルコフ網 (Hidden Markov Network: HMnet) を自動生成するアルゴリズムである。

HMnet は複数の状態を網状に結合したものとして表される。各状態に対して、それぞれ以下の情報を個別に与えることによって、その独立性が高められている。

- 1) 状態の番号
- 2) この状態で表現できる音素環境の集合
- 3) この状態に先行および後続する状態の集合
- 4) 出力確率分布のパラメータ
- 5) 後続する状態への切りかえタイミング

SSS は、ただ一つの状態から始めて、状態の分割を繰り返し行いながら HMnet の精密化を徐々に進めていく。SSS により、HMM モデルのパラメータと同時に、音素環境の分類も学習処理により決定される。

以下では、子音各音素について男女各 3 名の話者の音声データから SSS により生成された HMnet の分割状態を調べ、音声学的な考察を加える。

1. /h/

(a) 男性話者 15 分割

時間方向の前半部分に関して、/h-a/,/h-u/,/h-o/ の場合のみ語頭の /h/ が語中の /h/ と異なった軌跡を描いていた。つまり、後続音素が /a/,/u/,/o/ である /h/ は語頭、語中で異なった音素として認識されたと推測される。他の音素と比べて、時間方向の最前方の状態であるにも関わらず、後続音素の分割が目立つ。時間方向の真ん中に関して、後続音素の分割が目立つ。時間方向の後半部分に関して、/h-u/ が先行音素に関して細かく分割される。一般的に、他の音素に比べて、後続音素による分割が目立つ。時間方向の前半部分では、前舌母音 /i/,/y/,/e/ が同じグループに属し、語頭、語中の区別はない。/h-u/ は後半部分の時に先行音素に影響を受けて分割される。

(b) 女性話者 15 分割

時間方向の前半部分に関して、後続音素が /a/,/u/ である /h/ のみ語頭、語中の場合で異なった軌跡をたどる。時間方向の真ん中に関しては、後続音素の分割が目立つ。後半部分に関して、後続音素のみの分割になる。一般的に、他の音素と比べて、後続音素による分割が目立つ。

(c) 考察

日本語のハ行子音は後続音素にどの音素がくるかによって、異なった音声記号で表される。つまり、母音間の(特に /a/ に囲まれた) /h/ の場合、有声化の [h]- (つまり [ɦ])- になる傾向がある。(例として日本語の「母」-[ha-ha])。語頭の /h-u/ の場合、両唇無声摩擦音 - (つまり [ɸ])- に、/h-i/ の場合、硬口蓋無声摩擦音 - (つまり [ç])- になる傾向がある。SSS を用いた 15 の分割状態で、語中の /h-a/ が他の音素から分割したことから、他の /h/ と区別して認識されたと見える。又、時間方向の前半でその分割が行われたことから、[ɦ] を他の音素と時間方向の前半で区別して認識できたといえる。/h-u/ の語頭の分割状態から、時間方向の後半部分で /h-u/ を他の音素と区別して認識できたといえる。しかしながら、15 分割では、/h-i/ を他の音素と区別して認識していると観測することはできなかった。

2. /k/

(a) 男性話者 20 分割

時間方向の前半部分に関して、先行音素によって分割される。語頭、先行音素が促音とそれ以外の音素によって分割される。時間方向の真ん中では、先行、後続音素両方の影響を受ける。後続音素が /k-i/,/k-y/ の場合と後続音素が他の音素と区別される。後半に近い状態で、/k-a/,/k-u/ のみ先行音素の影響を受ける。後半では、後続音素の影響のみで、細かく分割される。/k-a/,/k-i/,/k-u/,/k-o/,(/k-j/ と /k-e/) これは、後続音素が前舌母音、後舌母音によって分割されている。一般的に、/k-a/,/k-u/,/k-o/ は語頭と先行音素が促音の場合とそれ以外の音素とに区別される。

(b) 女性話者 20 分割

時間方向の前半部分に関しては、先行音素、後続音素によって分割される。語頭、先行音素が促音の場合とそれ以外の音素によって分割される。時間方向の真ん中で

は、先行、後続音素両方の影響を受ける。先行音素が促音の場合、語頭の場合とそれ以外の音素に分けられる。後半部分では、後続音素の影響によってのみ、分割される。/k-a/,/k-o/,/k-u/,(/k-j/,/k-i/,/k-e/) 分割される。全般的に、/k-o/ は始めから他の音素と分割されている。先行音素が促音の場合と語頭の場合は、始めから分割されている。

(c) 考察

語頭、先行音素が促音とそれ以外の音素に分割される。これは、調音音声学的には促音が後続音素の /k/ の閉鎖部の延長として実現することを反映していると推測される。また、日本語のカ行子音は口蓋化した [kʲ] とそうでない [k] に区別される (前者は、/k-i/,/k-j/,(/k-e/) の子音、後者は、/k-a/,/k-u/,/k-o/ の子音)。すなわち、/k-i/,/k-j/ は後続音素 /i/ と /j/ (弁別素性 [-back,+high]) を共有しているといえる。前舌面が硬口蓋に向けて上昇するという特徴を共有しておりその調音面での特性が反映されたものと推測される。男性、女性とも時間方向の前半部分で口蓋化した [k] とそうでない [k] に区別される。しかしながら、女性話者の場合、口蓋化した [k] は同じ軌跡を描いているが、男性話者の場合は、/i/,/j/,/e/ がそれぞれ異なる軌跡を描いている。/k/ の場合、女性話者の口蓋化した [k] とそうでない [k] の違いは、男性話者のそれより区別して認識されやすいと推測される。

3. /m/,/n/,/ng/ ¹

(a) 男性話者 20 分割

時間方向の前半部分に関しては、先行音素、後続音素によって、分割される。時間方向の真ん中では、語頭の /m/,/n/,/ng/ (一か所) を除けば、後続音素のみで分割される。後半部分では、先行音素が /i/、語頭の場合を除けば、後続音素のみで分割される。全般的に、時間方向の前半、後半に比べて、真ん中での分割が少い。語頭の /m/,/n/,/ng/ は、時間方向の前半でそれら単独のグループに分割され、後半で /m/,/n/ と /ng/ に分割される。/m/,/n/ に関しては、後半で口蓋化した /mʲ/,/nʲ/ とそうでない /m/,/n/ に分割される。

(b) 女性話者 20 分割

時間方向の前半に関しては、先行音素、後続音素によって分割される。真ん中では、後続音素のみで分割される。後半部分では、語頭の /m,n/ を除けば、後続音素のみで分割される。全般的に、時間方向の前半、後半に比べて、真ん中では、分割が少い。語頭の /m/,/n/,/ng/ は時間方向の前半でそれら単独のグループに分割され、最後で、/m/,/n/ と /ng/ に分割される。

(c) 考察

日本語の「ン」とマ行の子音とナ行の子音を中心音素にして「ン」が時間の進行の中で、どちらの子音と近いのかということ調べてみた。結果として、時間方向の後半部分で、/ng/ と /n/,/m/ に分割されたが、/ng/ がどちらの音素に属するかということは観測することはできなかった。日本語の「ン」は、/p/,/b/,/m/ の

¹/ng/ は日本語の「ン」を示す。

前では、[m:]であり、/t/,/d/,/n/の前では、[n:]になる傾向があるが、30分割にして観測されなかった。(Nのサンプル数が多いという可能性と“duration”の違いに着目して分割された可能性があるとして推測される。)

4. /s/,/sh/²

(a) 男性話者 20 分割

時間方向の前半に関しては、先行音素のみで分割される。真ん中では、先行、後続音素の両方によって、分割される。ただし、語頭、先行音素が促音の場合は、他の音素の場合と区別される。後半では、後続音素のみで分割される。特に、後続音素が/i/,/j/の場合に限り、詳しく分割される。全般的に、/s/,/sh/が語頭にくる場合、時間方向の前半、真ん中で語中の場合と分かれている。後半では、後続音素が/i/,/j/の場合に限り、詳しく分割される。

(b) 女性話者 20 分割

時間方向の前半に関しては、先行音素によってのみ分割される。真ん中では、先行、後続音素の両方によって、分割される。後半では、後続音素によってのみ分割される。/sh/の後続音素が/i/,/j/,/u/のみ細かく分割されている。全般的に、/s,sh/が語頭にくる場合、時間方向の前半、真ん中で語中の場合と分かれている。後半では、/sh/だけが、中心音素の場合のみ細かく分割される。

(c) 考察

日本語のサ行は、「サ、ス、セ、ソ」の子音(音声記号では[s]で表される。)と「シ、シャ、シュ、ショ」(音声記号では[j]で表される。)は異なった音である。分析の結果として、男性話者の場合、時間方向の真ん中より後半部分から後続音素が/i/と/j/の場合の/s/は同じ軌跡を描いているが、女性話者の場合は同じ軌跡を描いていないことから、女性話者の場合、[j]を[s]と異なった音素と区別して認識されにくいと推測される。

5. /r/

(a) 男性話者 15 分割

時間方向の前半に関しては、先行音素によってのみ分割される。時間方向の真ん中、後半では、先行、後続音素によって分割される。全般的に語頭の/r/は他の場合と異なった音素であると推測される。後半で、/r-u/のみ先行音素によって分割されることがわかる。時間方向の前半部分で、/r/は細かく分割される。口蓋化した/r/とそうでない/r/を区別するのは、後半の部分である。

(b) 女性話者 15 分割

時間方向での分割の特徴は、男性と同じである。しかし、後半部分では、男性話者で観測された口蓋化した/r/とそれ以外の/r/は区別して認識されなかった。

(c) 考察

²/sh/は[j]を示す

一般に日本語の /r/ は破裂音の一種や弾き音として生起することが多く、時間軸の前方に破裂もしくは弾きに相当する部分と後方に後続音素への渡りの部分がある。時間軸の前半に分割された状態が、さらに語頭の /r/ と語中の /r/ に分割されているので、これは語頭の /r/ が破裂音の一種、語中の /r/ が弾き音として生起することが多いという調音的な特性に対応したものと考えられる。又、時間軸の後半に分割された状態が語頭の /r/ と語中の /r/ に分割されているのも同様の理由によるものである。/j/ の先行音素が /r/ の場合、他の子音が先行音素の場合に比べて /j/ の部分が長くなる傾向があるといわれている。しかし、/r/ の後続音素 /j/ は 15 分割では、この特徴を裏付ける分割状態に至っていない。(口蓋化した /r/ という点で、/i/ と同じ軌跡を描いている。)

4 おわりに

本研究は、ATR 自動翻訳電話研究所で行っている音素環境を考慮した高精度な音声認識の手法として開発された「逐次分割法 (SSS Algorithm)」を音声分析に応用して、その結果得られた隠れマルコフ網のネットワーク状に結合した複数の音韻環境の状態を、音声学的な観点から検討を試みた研究成果の途中報告である。まだ分析途上の段階であるが、音素の異音の分析状況や、従来のパラトグラフィーや X 線トレース図などによる研究によって進められている調音機構の変化の状態や共調音 (co-articulation) 効果の現象を音響的な側面から解明する糸口が得られた。今後は、さらにサンプル数を増やす共に分析条件を整えて研究を進めていきたいと思う。

謝辞

本研究の機会を与えて下さった ATR 自動翻訳電話研究所博松社長、並びに御助言いただいた、同研究所の研究員の皆様方に深く感謝致します。

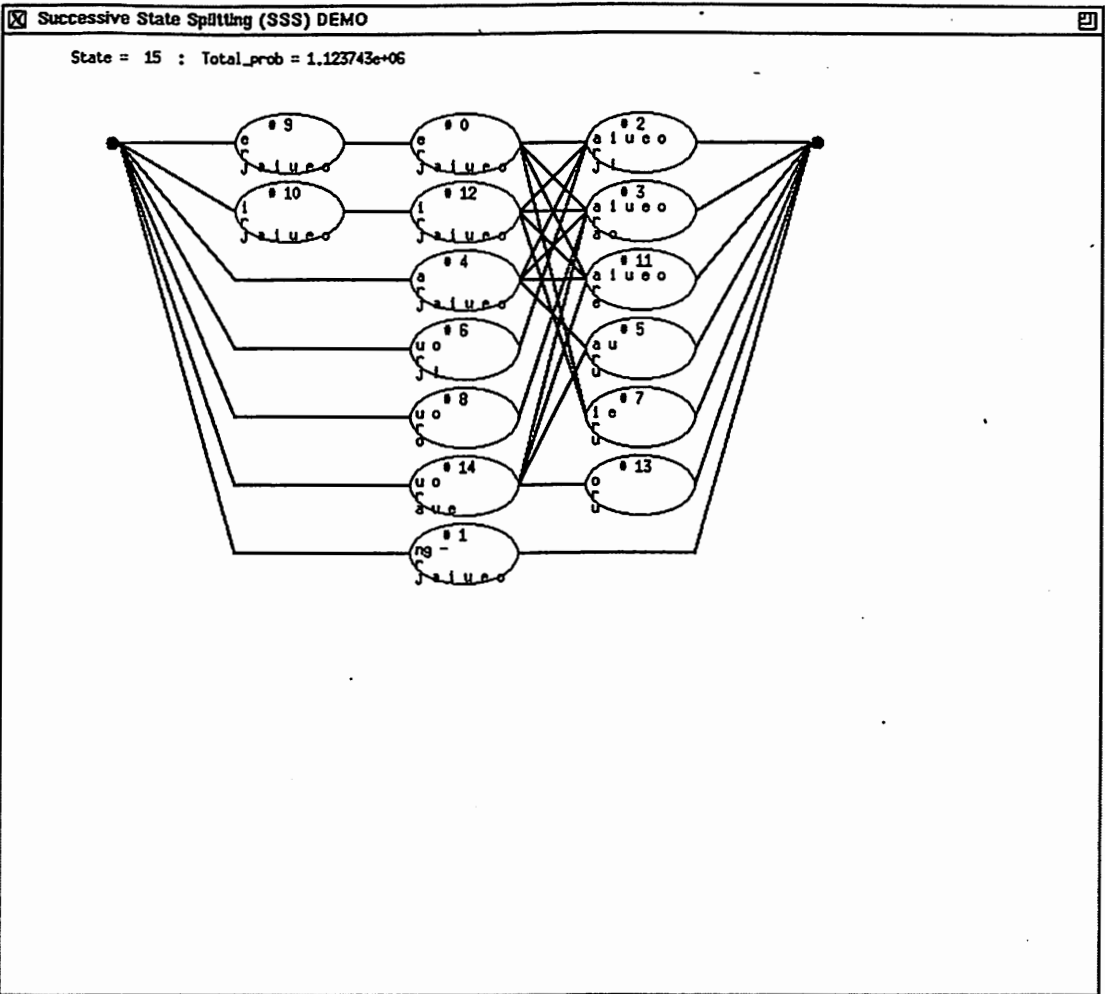
参考文献

- [天沼, 他 1986] 天沼寧, 大坪一夫, 水谷修, 『日本語音声学』 (くろしお出版), 1986
- [壇辻 1992] 壇辻正剛「弁別素性の音響的側面に関する一考察」(『関西大学視聴覚教育』第 15 号、pp.42-74) 1992
- [鷹見 1992] 鷹見淳一, 永井明人, 嵯峨山茂樹「aka と aki は同じ音? - 前後の音素を考慮した高精度音声認識 -」(『ATR ジャーナル』1992 秋, No 12) 1992
- [鷹見 1991] 鷹見淳一, 嵯峨山茂樹「音素コンテキストと時間に関する逐次状態分割による隠れマルコフ網の自動生成」(『電子情報通信学会技術研究報告』SP91-88) 1991
- [壇辻 1989] 壇辻正剛, 嵯峨山茂樹「音素環境クラスタリングと弁別素性の音響的側面に関する一検討」(『電子情報通信学会技術研究報告』) 1989

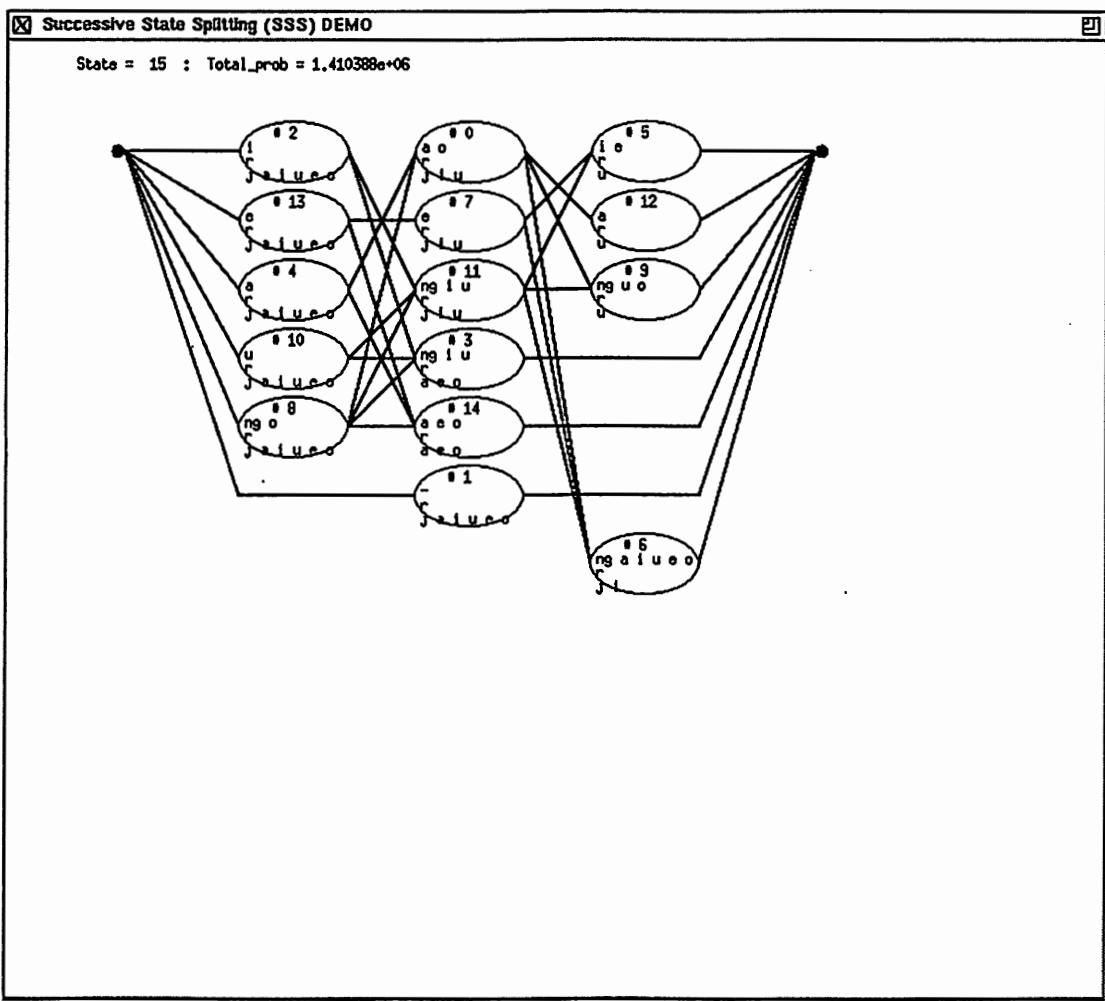
[Dantsuji and Sagayama 1991] Dantsuji,M. and Sagayama,S “ A Study on Distinctive Features and Feature Hierarchies through 'Phoneme Environment Clustering' (PEC)” (『Proc. Xllth Int. Congress of Phonetic Sciences』 Vol.3, pp.190-193) 1991

[壇辻 1992] 壇辻正剛「音声. 音韻. 韻律」(『日本語を学ぶ人のために』), 世界思想社 1992

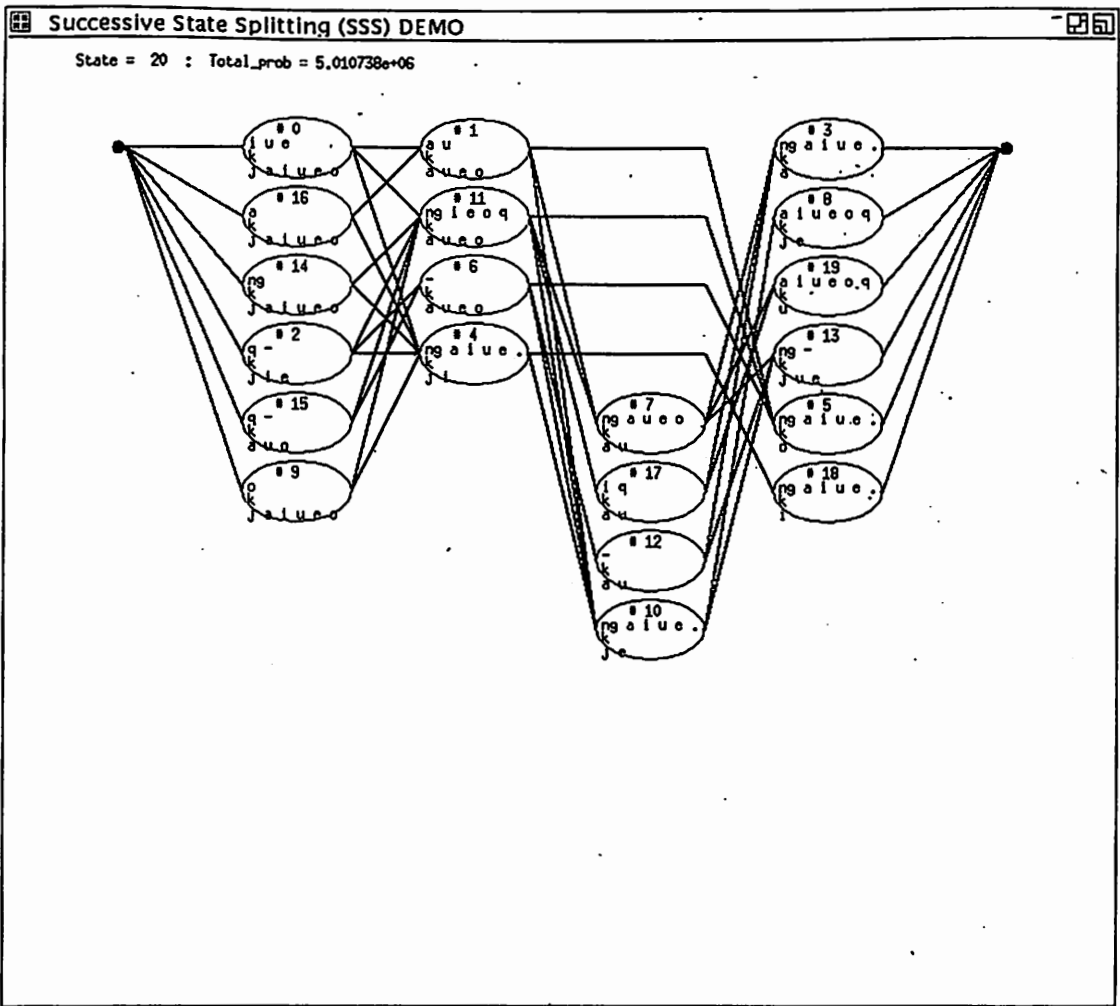
付録 “SSS” による日本語子音の分割表



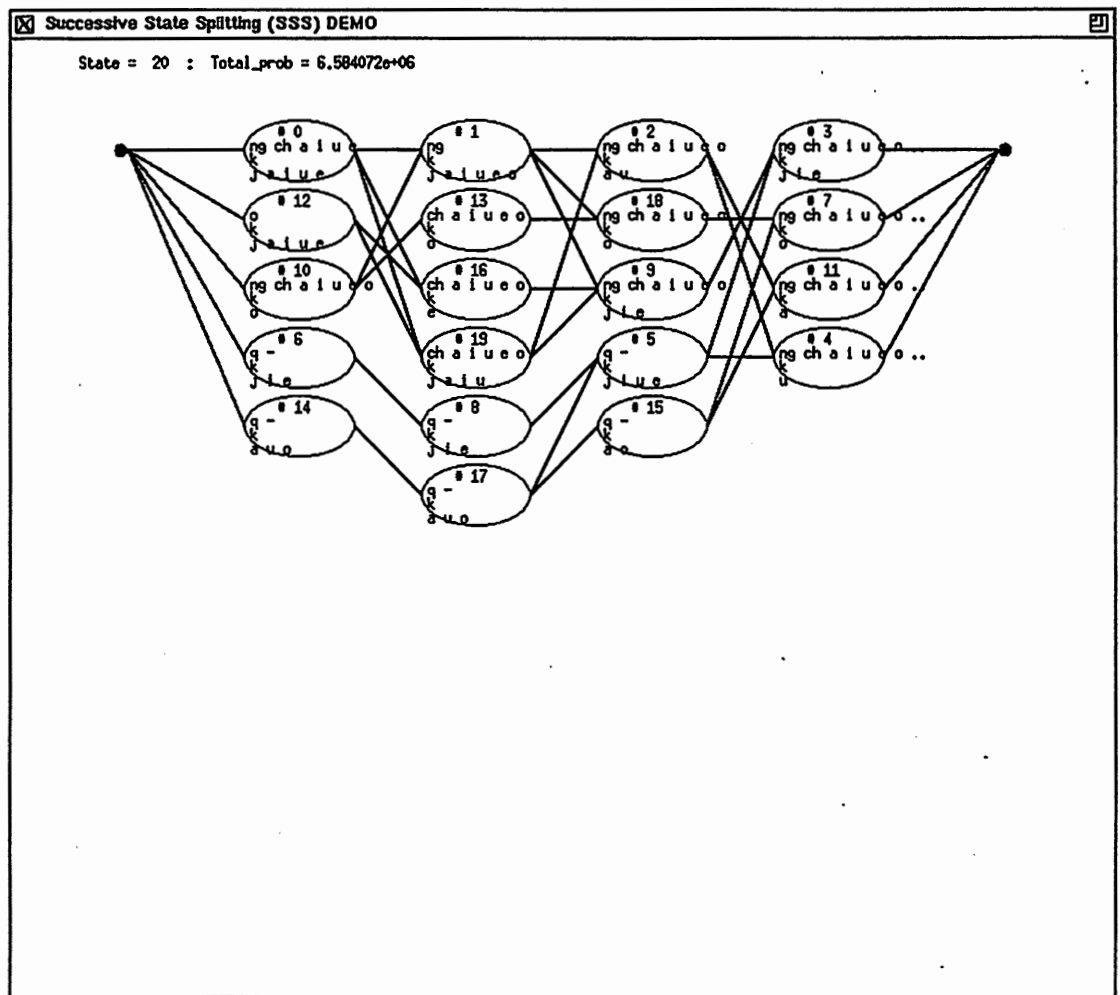
1/1
男性話者
15分割



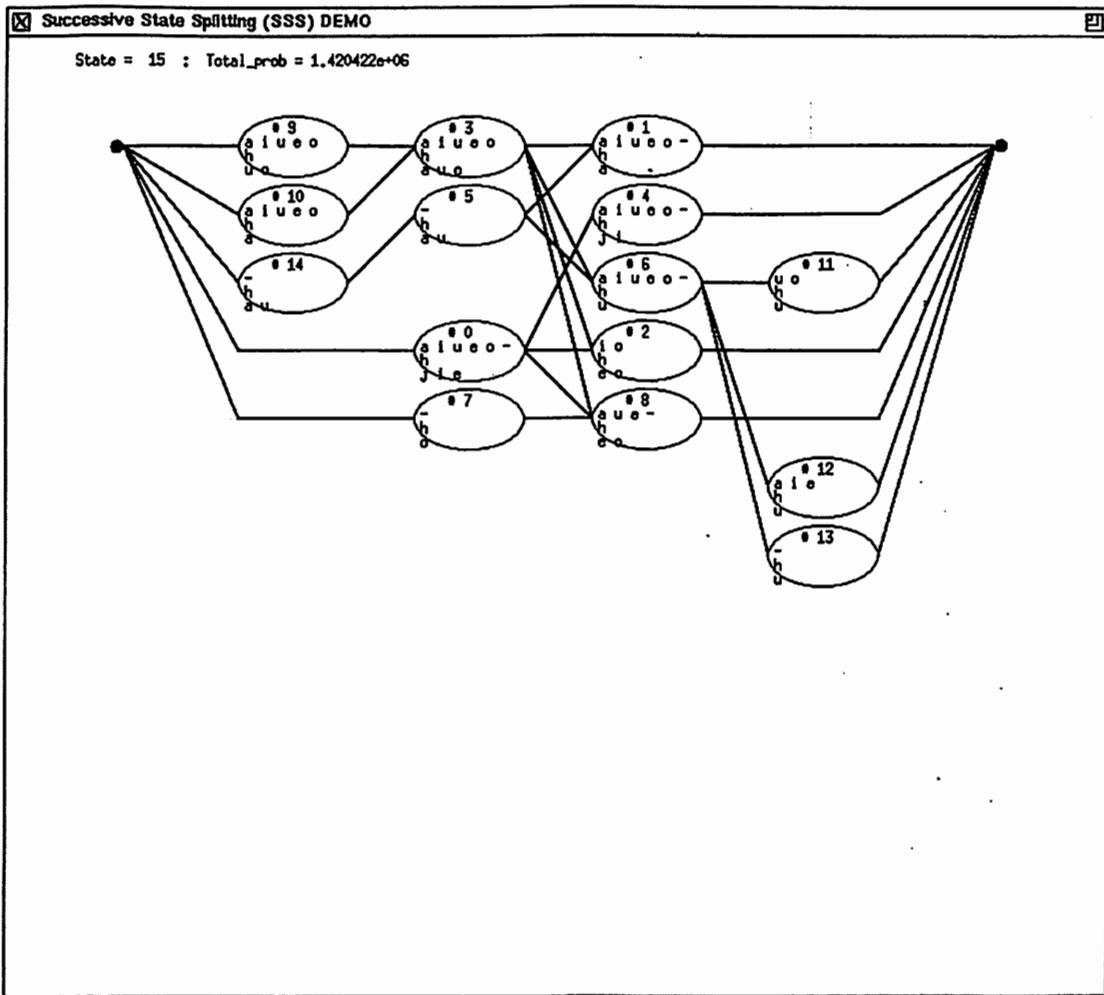
1/1
女性話者
15分割



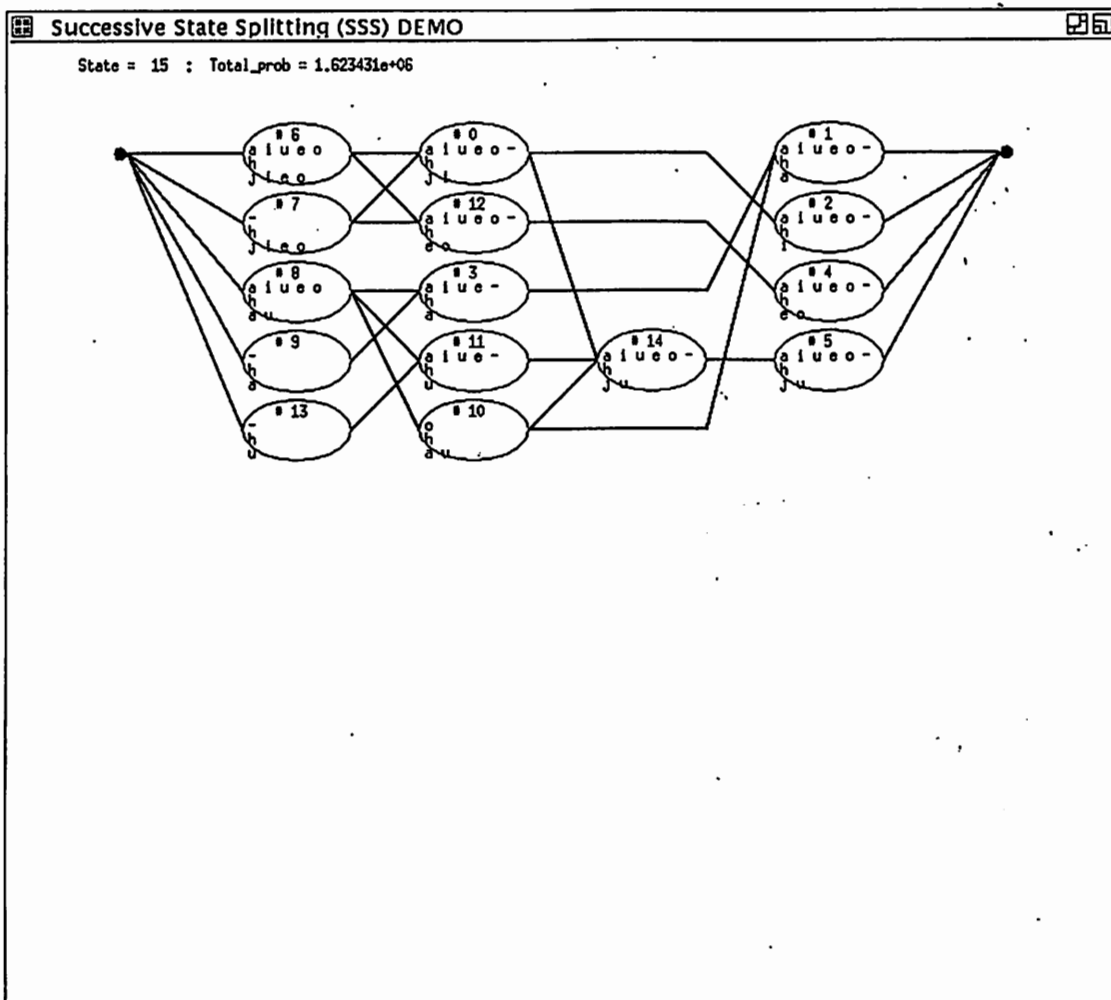
1/k/
男性話者
20分割



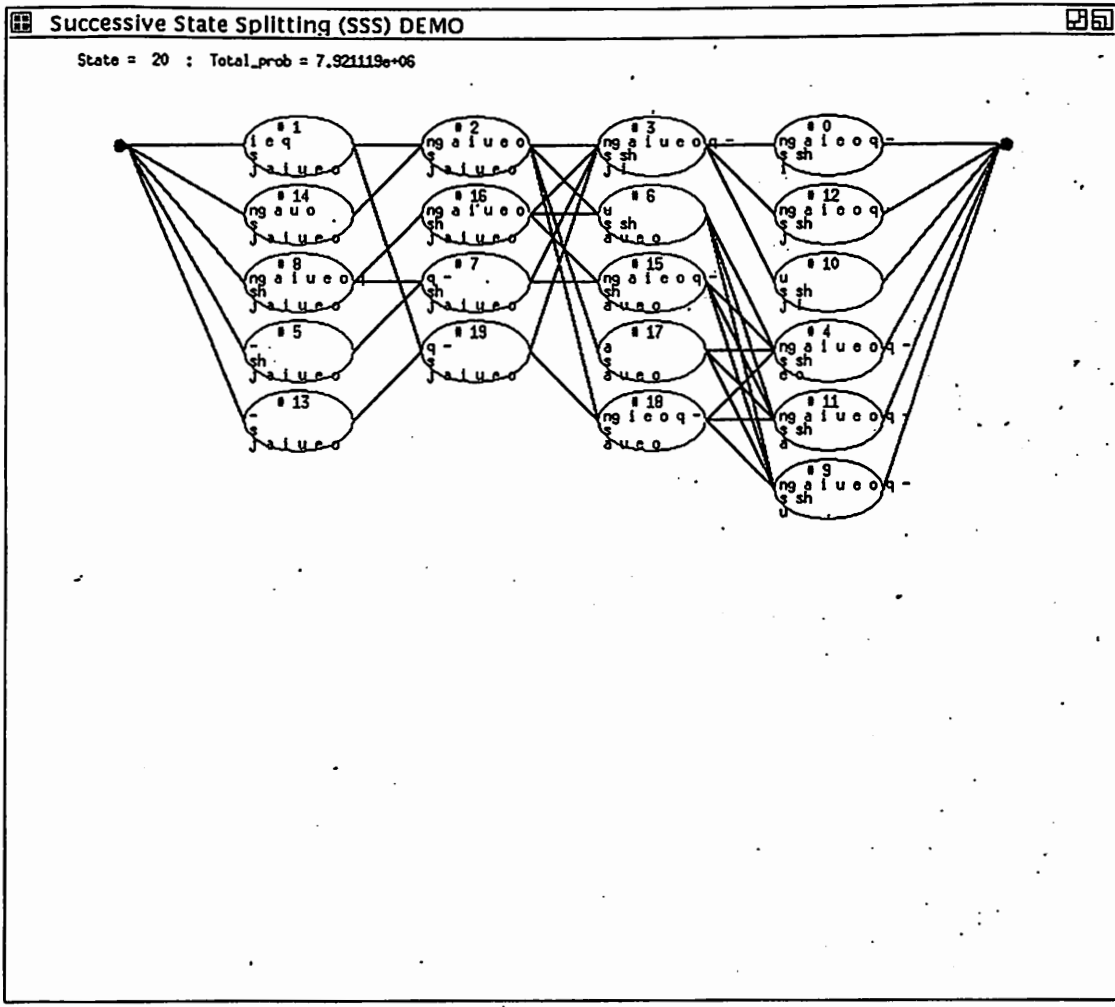
1/k/
女性話者
20分割



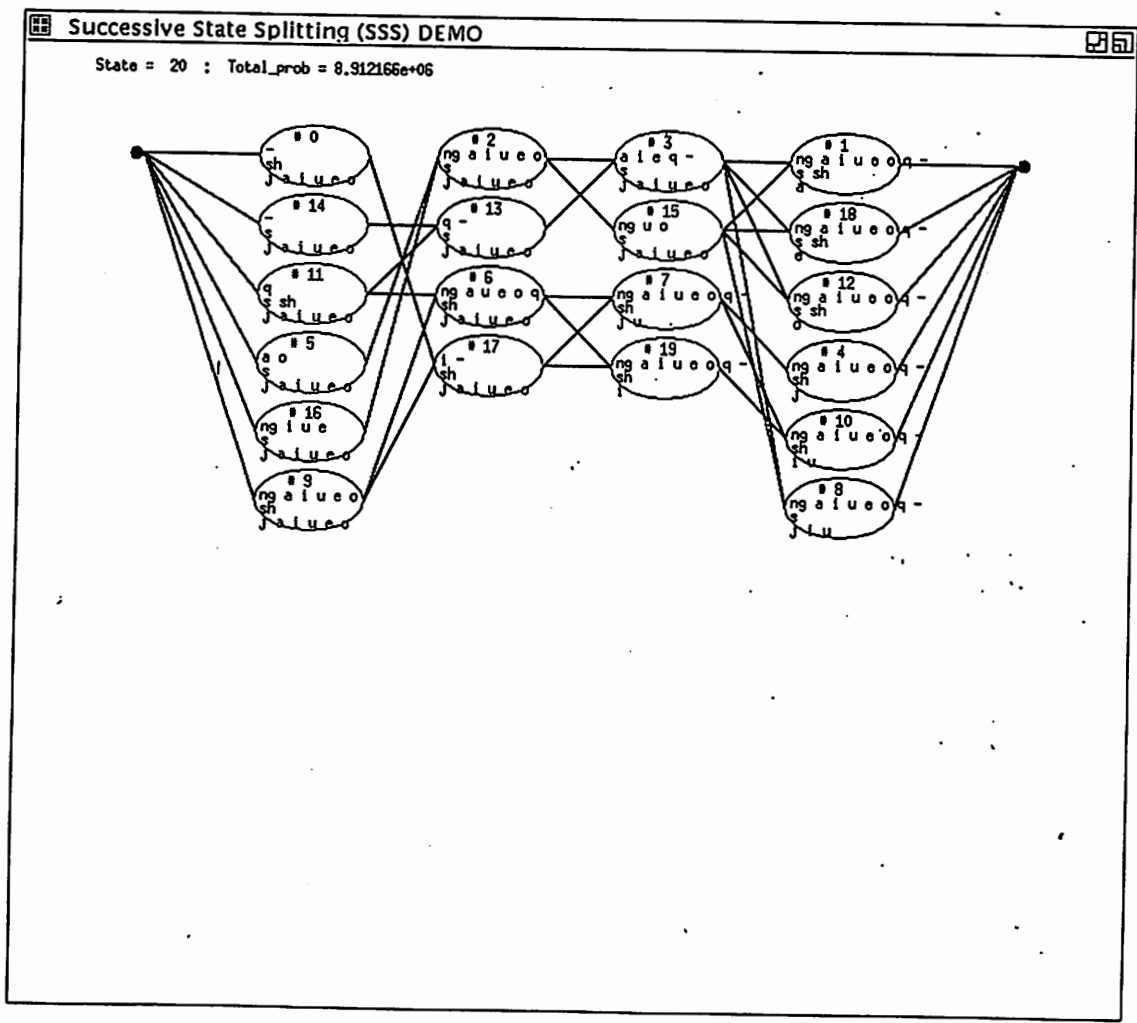
/h/
 男性話者
 15分割



/h/
 女性話者
 15分割



1S, 5h/
 男性話者
 20分割



1S, 5h/
 女性話者
 20分割

