講演調の話し言葉に対する分析

峯松 信明　片岡 嘉孝　中川 聖一
mine@*domain*　kataoka@*domain*　nakagawa@*domain*
domain = slp.tutics.tut.ac.jp

豊橋技術科学大学 情報工学系
〒441 愛知県豊橋市天伯町雲雀ヶ丘 1-1

あらまし　本研究では講演調の話し言葉に対して，音響的/言語的，更には知覚的観点から分析を行なった。特に，講演調の話し言葉に対して人間が感じる「ポーズ(間，区切り)」が音響的(物理的)なポーズとどの程度対応がとれるのか，そして，知覚的ポーズの周辺にはどのような言語表現(間投詞，つなぎ語，終助詞)が頻出するのか，の2点に焦点を置いた分析を行なった。その結果，音響的ポーズと知覚的ポーズとの相関には発話速度が関与していることが示された。また，知覚的ポーズをほぼ確実に引き起こす言語表現として「え[~]」「え[~]と」「で」が観測された。なお本報告では，50年代より言語学者らによって行なわれてきた話し言葉に対する研究例のサーベイも行なっている。これらの研究例を考察することは工学的応用と言う観点から考えた場合においても，非常に有益なことである。

和文キーワード　話し言葉，講演調，音響的/知覚的ポーズ，間投詞，つなぎ語

Analysis of Spoken Language in Lecture Style

Nobuaki MINEMATSU　*Yoshiyuki KATAOKA*　*Seiichi NAKAGAWA*
mine@*domain*　kataoka@*domain*　nakagawa@*domain*
domain = slp.tutics.tut.ac.jp

Department of Information and Computer Sciences, Toyohashi University of Technology
1-1 Hibarigaoka, Tenpaku-chou, Toyohashi-shi, Aichi, 441, JAPAN

Abstract　Analysis of spoken language in lecture style was carried out from acoustic, linguistic and perceptual viewpoints. Especially, the correlation was investigated between pauses which human listeners perceive in lecture-style speech and those which were detected semi-automatically using some acoustic methods. Linguistic expressions(interjections and filled pauses) around the perceptual pauses were also analyzed. As a result, it was found that the correlation between the two types of pauses was influenced by speech rate and that "e[e]", "e[e]to" and "de" were observed as the linguistic expressions which caused perceptual pauses in quite high probability. And in this paper, some of the traditional researches for spoken language conducted not by engineers but by linguists were also surveyed. It is very beneficial to look into these researches in terms of technological application.

英文 key words　spoken language, lecture-style, acoustic/perceptual pause, interjection, filled pause

1 はじめに

計算機技術の進展に支えられ，音声認識(技術)が対象とする音声は，単語から文，そして対話に至るようになった。対話音声を対象とした場合，そこには“書き言葉”には見られない，“話し言葉”特有の言語現象(言い淀み，言い直し，倒置，省略，間投詞/つなぎ語の使用，“こそあど”語の使用，etc)が頻出する[1]。最近，話し言葉の中でも焦点を対話音声に絞り，その特有の現象を分析した研究が行なわれているが[2]-[5]，話し言葉と書き言葉の中間的な性質を持つと考えられる講演調(独話形式)の言語現象に対しては，その分析例が少ない。

そこで本研究では，対象を講演調の話し言葉に限定し，人間が感じるポーズ(間，区切り)が音響的ポーズとどの程度相関を持つのか，と言う観点から分析を行なった。また，知覚的ポーズを引き起こし易い言語表現の存在も考えられる。そこで，知覚的ポーズ周辺の言語表現(間投詞，つなぎ語，終助詞)に着目した分析も行なった。

一方，言語学者らによる話し言葉に対する分析は，日本においても50年代半ばより精力的に行なわれている[6]-[13]。本報告では，これら言語学者らによって行なわれてきた種々の研究についても，そのサーベイを述べることにする。言語学者らによるこれらの研究は，定性的な分析に終始するものが多いが，音声対話を工学的に扱う場合においても，有益な情報を与えるものである。

2 本研究の背景と目的

音声認識の対象が，文/対話音声になるにつれ，話し言葉における種々の言語現象を定量的に分析した研究が広く行なわれている。一方，最近，書き言葉用の文法に基付く規則を，音声認識用の文法として利用することに対する是非が問われている。1995年5月に行なわれた音声情報処理研究会/自然言語処理研究会パネルディスカッションにおいても，「話し言葉の文法構築は可能か？」と言うタイトルの下，議論が交わされた[14]。その際，ポーズに基付く文法記述，即ち，ポーズ～ポーズ間を一つの処理単位とする文法記述が一つのトピックとなり，その可能性についての討論も行なわれた[4][5][15][16]。

従来の音声認識では，ポーズに対応する(無音の)音韻モデルを作成するなどの処理のみが行なわれていたが，このポーズを積極的に利用しようと言うものである。書き言葉では(一般的に)表現されない，話し言葉のみに存在する要素を積極的に利用する方向性は非常に興味惹かれるところである。

しかし，積極的にポーズを利用する場合，その定義が問題視されるべきであると考えられる。従来，ポーズとは「ある閾値以下(かつ，ある閾値以上の時間長)の音声(*)」を意味していた。しかし，このポーズを利用して入力音声の処理単位を定義することを考える場合，人間の所作を考慮する必要が生じる。英語を母国語とする人にとっては，「ハト」も「ハット」もどちらも“*hat*”であり，両者の区別は非常に困難であるとの話をよく聞く。これは，促音(無音が一つの音韻となる)現象が少なくとも英語には非常に稀であることを示している。このように，言語が異なれば(音響的に)同一のポーズが異なって知覚されるように，同一ポーズであっても，前後のコンテキストによっては異なった解釈がされることが考えられる。人間が感じとるポーズ(「間」と呼ぶべきかもしれない)と音響的に(例えば(*)で)定義されるポーズとがどの程度相関を持っているかを調べることは，純粋に音声知覚的な意味での重要性の他に，今後ポーズを積極的に利用した音声処理手法を構築する際においても，十分に考慮されるべきである。

本報告では以下，音響的に(例えば(*)で)定義されるポーズを**音響的ポーズ**と呼び，人間が「間」として知覚したポーズを**知覚的ポーズ**と呼ぶことにする。なお，知覚的ポーズ前後のコンテキストを言語的に分析するに当たっては，間投詞/つなぎ語/終助詞の存在を考慮する。これは音声のある部分(主に音響的ポーズ部分)が「間」として知覚され，(音声が区切られ)処理単位となっているのか，上記のような言葉表現がマーカーとして働き，そのマーカーの知覚が先立って行なわれ，(音声が区切られ)処理単位となっているのか，を検討することを目的としている。

第1節で述べたように，本研究では比較的例の少ない，講演調(独話形式)の音声資料を用いる。これは従来行なわれてきた対話音声における分析例[3]などと比較することによって，間接的にではあるが講演調の話し言葉の特徴を明らかにすることを本研究の目的の一つとしているからである。また，対話音声では1 turnの発話長が比較的短く，話し言葉の音響的特徴を調べる上では講演調の音声が適している面もある。

なお，話し言葉に対する研究は国語研究所を中心に50年代半ばより精力的に行なわれている。特に話し言葉の文型(sentence pattern)の分析においては[7][8]，文字化したデータ以外に，イントネーションを考慮した上で，文型の範疇化を試みている。当時の分析環境，及び当事者の音声/音響学的な知識不足が影響し[7][8]，その分析結果は定性的なものに留まっているが，研究の方向性としては現在の音声言語情報処理が向かうべき方向性を示唆しているようにも思える。そこで本報告では，これら言語学者らによる研究例を次節で簡単に紹介し，その後で，筆者らが行なった分析及びその結果について述べることにする。

3 言語学者らによる話し言葉の分析

1950年代から60年代にかけて，国立国語研究所において，大がかりな話し言葉の収録と分析が行なわれている[6]-[8]。「**談話語の実態**」[6]においては，録音した日常の談話を書き起こし，1) イントネーション，2) 語・文節・文の長さ，3) 文の構造，4) 語の種類・使用度数・用法，など，言語の基本的な性質を統計的に求めている。この分析に当たって使用された音声資料は，日常談話の他，その比較資料としてニュース(解説)・座談会・楽語・講義など幅広く資料収集を行ない，かつ，発話者においても“女子学生”から“じいさん・ばあさん”まで幅広く収録されている。

そして，得られた分析結果を書き言葉である新聞からの分析結果と比較することで話し言葉の特徴を述べている。例えば・文節数・文の構造・語の種類・使用度数から日常談話と新聞文章とを比べると，日常談話では・1,2 文節文が全体の半数を占める・主語のある文が少ない(26%)・倒置文が多い(7%)・名詞が少ない・名詞と動詞の比較において，動詞が多い・動詞と形容詞の比較において形容詞が多い・コソアド語が多い・文末助詞を有するものが73%で，そのうち「ね(25%)」「よ(15%)」「の(7%)」「か(6%)」が目立つ・間投詞の付いている割合は，主語は8%，連体修飾語は3%，連用修飾語は12%である，などの特徴を上げている。しかしこの分析において，「文」の定義はなされておらず，話し言葉を「文」として書き起こすことができることが前提となっているようである。

一方「話し言葉の文型(I)−対話資料による研究−」[7]及び「話し言葉の文型(II)−独話資料による研究−」[8]では，まず，「文の認定」から作業が開始される。即ち，この分析において文と認める「発話」と文として不完全な「発話」(例えば，「お名前は？」などの省略文も不完全文となる)とを区別し，文と認められる「発話」のみを扱う，とするものである。このように文の定義を行ない，(I)では対話音声を(II)では独話音声(講演，講義，演説，解説)の分析を行なっている。しかし，各々の分析結果からの，対話と独話の詳細な比較は行なわれておらず，比較を目的とした分析ではないとも明示されている。更に，不整形と言われる間投詞，助詞落ち，倒置，言い直し，言い淀みなどの分析も行なっていない。両者の分析方針はほぼ同じであり(当然(I)から(II)へと改良された点もあるが，ここでは触れない)，以下の通りである。即ち，文型(sentence pattern)を1) 表現意図，2) 構文の型，3) イントネーションの型の総合として捉えようと言うものである。ここで表現意図とは，文字面から直接的に受けとれる意図を意味し，大きく・詠嘆表現・判叙表現・要求表現・応答表現に分類している(tree 状に更に細かく分れる)。そして，各種表現意図と主に文末の表現形式との対応を調査している。構文分析においては，文末述語を中心として，それと直接関係を結ぶ構成要素の配列・組合せによって構文の型を調査している。イントネーションにおいても同様に，文末に限定した分析が行なわれている。文末部のイントネーションを・平調・昇調1・昇調2・降調・特殊型に分類し，各種イントネーションと表現意図との関係を調べている。以上の3つの観点からの分析を基に，最終的には表現意図に対応する各種の文(末)表現において，どのような構文の型，イントネーションの型が用いられるかを整理している。しかしながら，純粋に音声要素であるイントネーションの扱いには終始苦労していたようである。実験への反省として，音声要素に対する(技術的)知識不足が大きく取り上げられている。今回筆者らが調べた限りにおいては，以上の3つの文献が実際に「話し言葉」を収録・分析・考察した形の研究となっている。いずれにおいても，定性的な考察が多いが，録音機器の取り扱いにも十分な注意が必要であった当時に，これだけのデータを収集し，分析したと言う事実は驚嘆すべきものがあると思う。

次に，内省的な考察により話し言葉の諸現象を考察した研究例について述べる。「はなしことば序」[9]において伊佐早は，不整表現を生む原因を心理的/認知科学的な面から考察している。即ち，発話とは，無限の可能性のある心の動きを，無限に近い制約のあるコトバに変換することであり，しかも，やりなおしの利かない時間の上に，一瞬の選択により，コトバを並べる作業であると言っている。更に，発話されたコトバに対する配慮の乏しさを指摘し，その結果，話し言葉の構成要素間の粘着力は非常に低いものになる，としている。換言すれば，非常に狭い範囲においてのみ「筋が通る」場合が頻出し，論理の息が短くなってしまう，のである。この論理をより広範囲において成立させようとした場合，空間的な推敲は不可能であるから，結局は新たな時間の上にコトバを言い足し言い足しすることになってしまう。その結果，重複表現が多発することにもなるが，この重複表現については，話し言葉の整理されない面を表すと共に，一方では注釈，強調，口調を整えるのに役立っている，と述べている。また，発話の区切れに対しても考察が行なわれている。即ち，発話の区切れにおける特徴として・あゆひ(文末の助詞，助動詞など)・尾高イントネーション・間(ポーズ)・頭高イントネーション・かざし(文頭の間投詞，接続詞，副詞など)を上げており，これらの音声/言語現象を用いて，話しの内容を色付けて話し手の態度を表すと共に，相手に対する働きかけをしている，と述べている。

文献[7][8]の中心的研究者である大石は，「話し言葉とは何か」[11]において，話し言葉と書き言葉の比較を行なっているが，ポーズに関して次のように述べている。即ち「日常の話し言葉の上では，ポーズは，その置かれる場所に関しても，大きさの別に関しても，一般に極めて不安定で，なんの意味をも持たないものが多い。しかし，右に挙げてきたような文法的構造を明らかにするものから強調その他特殊表現を担うポーズがあるので，これを軽視するわけにはいかない。」

「はなしことばにおける文法」[12]において石井は，書き言の文に対応するものとして，発話交替のタイミングを上げている。対話が成立するためには，適切なタイミングで両者が互いに発話する必要があり，そのタイミングが書き言葉における「文」に対応するものである，としている。そして，書き言葉の句点に相当するものとして，ポーズ，イントネーション，言語形式，意味などが関与すると述べている。更に，見ぶり手振りなどの視覚情報の不可欠さについても考察している。

「話しことばの文法」[10]において宇野は，話し言葉の特徴を述べると共に，「話し言葉」と言う用語の定義を明確化している(但し，この論文の中においての話である)。即ち，会話・談話風の文体を持つことばを「話しことば」とし，音声を媒体とすることばを「音声言語」とし，次のような分類を行なっている。

1. 話しことば・音声言語 → 例えば，談話
2. 話しことば・文字言語 → 例えば，会話文
3. 書きことば・音声言語 → 例えば，朗読
4. 書きことば・文字言語 → 例えば，文章

4 講演調の話し言葉の音響的分析

第2節でも述べたように，講演調の話し言葉における音響的/言語的及び知覚的分析を行なったので，本節でまず音響的分析について述べる。

4.1 音声資料

本分析で用いた音声資料は，1995年5月に行なわれた音声情報処理研究会/自然言語処理研究会パネルディスカッション(「話し言葉の文法構築は可能か?」)における前半部分，即ち5人のパネラー(以降，**SP1～SP5**と記す)が，各自の主張を順々に述べている部分である。なお，このパネルの模様は，後半のディスカッション部分も含めて，

http://www.slp.tutics.tut.ac.jp/SLP/panel-May-95.html

にて公開中である。パネルの内容についてはこちらを参照して頂きたい。但し，公開してあるのは，各発言者による修正済みのものであり，討論の様子を正確に書き起こしたものではない。**表1**に使用した音声資料における，発話者別の音声長/モーラ数/発話速度を記す。但しモーラ数は，書き起こしテキストに基付いて算出したものであり，長音などの扱いは，テキスト化の段階での書き起こし者の主観に依るものであることを断っておく。また，音声長は，講演開始から講演終了までの全時間を意味しており，無音部の除去等の処理は行なっていない(無音区間を考慮した考察は**第6.1節**で述べる)。

なお，講演(独話)形式の音声は一般に，話し言葉と書き言葉の中間に位置すると言われるが，今回扱った音声資料は，講演内容の(読み上げ)原稿を予め作成しておくと言ったことはなく，講演中に使用されたのは，OHP数枚程度である。

表1. 各話者別の音声長/モーラ数/発話速度

	音声長 [sec]	モーラ数	発話速度 [mora/sec]
SP1	741.5	5153	6.95
SP2	570.3	3726	6.53
SP3	680.1	4915	7.23
SP4	351.5	2380	6.77
SP5	518.5	4184	8.07

4.2 知覚的ポーズの検出及び定義

上記の音声資料に対して，人間が感じ取るポーズ(間，区切り)の検出を以下のようにして行なった。成人男性6人，成人女性1人の計7人を被験者として実験を行なった。被験者には，正確に書き起こした資料を配布し，「音声聴取時に，ポーズ/間/区切り，を感じた所に鉛筆で"／"印を記入する」よう指示した。この際「長音化された部分に引きつられて"／"を入れぬよう」注意を促した。また，練習用として冒頭の司会者の発言(約1分)を使ってポーズ検出を行なわせ，予め用意しておいたポーズ検出例と参照させることにより，被験態度の統一を図った。なお，実際のポーズ検出に当っては，書き起こし資料1頁(約70秒に相当する)毎にテープを止め，再度その頁の音声を提示し，ポーズ検出の確認・チェックを行なわせた(計2度の聴取)。

各話者に対するポーズ節数は，当然のことながら被験者によって異なる。そこで，得られた結果より，各話者毎の平均ポーズ節数を求めると同時に，被験者のうちn人以上(n=1～7)がポーズが存在すると判断した箇所(知覚的ポーズの候補時点)を集計した。知覚的ポーズの位置/数は，このnに依存した形で選定されることになるが，その数が実験より求まったポーズ節数の平均値に最も近くなるnを選ぶことで，知覚的ポーズを定義した。その結果n=4となった。被験者数が7であることを考えると，過半数の人間がポーズがあると判断した時点を本実験における知覚的ポーズとして定義することとなり，この定義の妥当性が示唆される。なお**表2**に各話者に対する7人の被験者が示したポーズ節数の平均値と，上記の通り定義された知覚的ポーズ節数(n=4)を示す。両者には，殆んど差が無いことが分る。

表2. 各話者別の平均ポーズ及び知覚的ポーズ節数

	平均ポーズ節数	知覚的ポーズ節数
SP1	297.3	299
SP2	269.0	266
SP3	302.6	296
SP4	151.3	142
SP5	218.6	197

4.3 音響的ポーズの定義及び検出

音響的な情報のみに基付いてポーズを検出した。但し，今回利用した情報はパワーのみである。ポーズ節についての分析例[3][4][16]を見ると，現在のところ，ポーズの検出については人手を介しているようである。しかし，実際の応用を考えた場合，全てを自動化すべきであるのは言うまでもない。そこで本研究では，以下のように半自動化して音響的ポーズ検出を行なった。まず音声区間を

閾値θ以上のパワーが時間τ以上継続した区間

として定義し，それ以外を無音，即ちポーズと定義する。ここでθ，τは発話者/使用したマイク等に依存するパラメータである。そこで，各話者の講演音声を頭から60秒毎に区切り，**第4.2節**で定義した知覚的ポーズが最も多く含まれる区間を話者毎に一つずつ選ぶ(平均ポーズ数29.6個)。選ばれた区間に対しては，各知覚的ポーズの開始・終了時点を正確に波形視察・音声聴取に基付いて決定する。それ以外の区間に対しては，知覚的ポーズの(ほぼ)中心点を波形視察・音声聴取により求めておく。次に，選ばれた区間に対して，上記の方法でポーズ検出を行ない，知覚的

ポーズ区間の検出誤差が最小になるよう各話者毎にθ, τを求める。求めたθ, τを用いて各話者の講演音声全体を，上記の方法によりポーズ検出し，これを音響的ポーズと定義する。これより，各知覚的ポーズの中心時点が含まれる音響的ポーズを検索すれば，非選択区間における知覚的ポーズの開始・終了点も半自動的に求まることになる。なお本研究では，促音或は破裂音前の無音区間の除去などの操作は行なっていない。これは音声処理単位の決定にポーズを導入する場合，ポーズ検出は認識処理の前処理として位置すべきであると考えられるからである。

音響的ポーズと知覚的ポーズの含有関係を見ると，小数の例外的な場合を除いて，"知覚的ポーズ$\in$音響的ポーズ"となっていた。例外的な場合とは，波形視察では無音部が存在しない区間に対して，ポーズがあると判断された場合を言い，音響的以外の要素が原因となっているものと思われる。以降の考察では，この例外的な場合は除いて考えることにする。

図1(a)～(e)に各話者に対する音響的/知覚的ポーズの度数分布を示す。横軸のポーズ長は50[msec]単位で離散化してある。なお，図では50[msec]以上のポーズのみを対象とし，50以上の度数は50として表示している(50～100[msec]の度数は200～400程度である)。更に，これらを「d[msec]以上の音響的ポーズが知覚的ポーズとして判断される割合(確率，†と記す)」と言う観点から描き直したものを図2に示す。横軸がポーズの下限長dを表している。但しdの増加と共に，該当する音響的ポーズの総数が減少し，算出される確率の信頼性が乏しくなる。そこで各話者において，d[msec]以上の音響的ポーズが$N(=20)$個以上ある場合のみをプロットした。なお，dの増加と共に†が減少している話者がある。これは知覚実験において，長い音響的ポーズが「区切り」としてマークされなかったことを意味している。原因としては，OHP操作などで生じた(即ち非言語的活動で生じた)長い音響的ポーズに対する被験態度の違いや，音響的ポーズそのものの自動検出エラー(長いポーズが「音響的ポーズ+短時間の音声+音響的ポーズ」と誤認識されてしまう)が影響しており，この点の処置は今後の課題の一つである。なお，音響的/知覚的ポーズ間の相関については**第6節**で考察する。

5 講演調の話し言葉に対する言語的分析

第2節でも述べたように，間投詞/つなぎ語/終助詞と言う観点から，講演調音声データの言語的分析を行なった。

5.1 間投詞

表3に今回取り上げた間投詞の種類と，各話者別の出現頻度を示す。なお"[X]"とは，"X"が省略可能であることを示す。どの話者にも共通して用いられるもの(ま[～])もあるが，特定の話者のみに用いられるもの(あの[～]，え[～]と，等)もある。後者は発話者の発話スタイルを表現しているものと考えられる。また，間投詞全体の出現数を見ると**SP4**が飛び抜けて出現頻度が高いことが分る。知覚的ポーズとの関連については**第6節**で考察する。

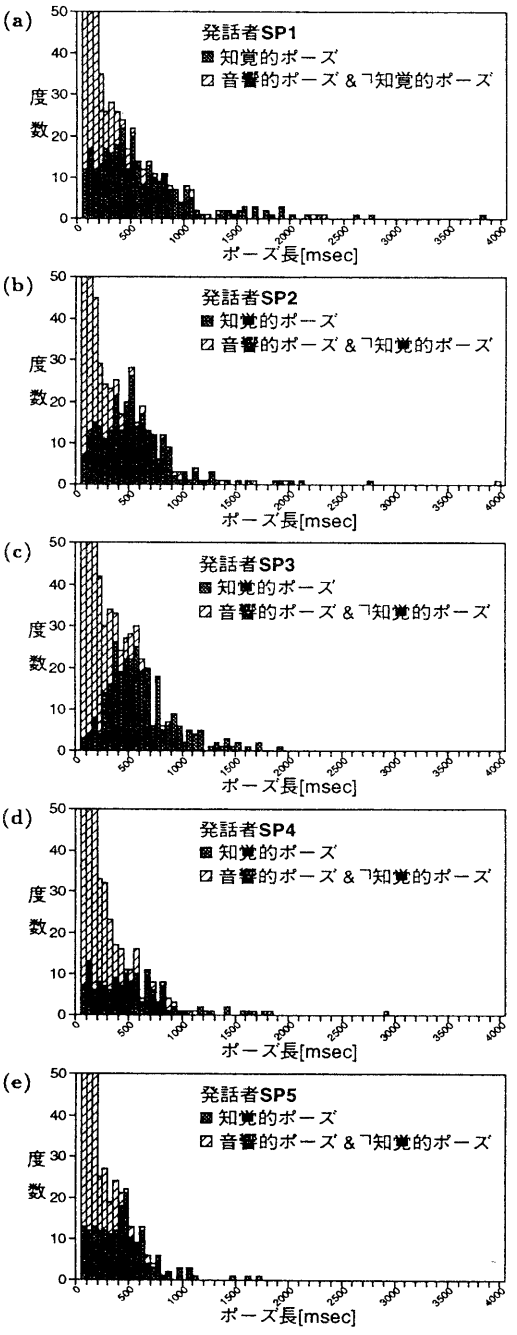

図 1. 各話者における音響的・知覚的ポーズ
(a)～(e)は**SP1**～**SP5**に相当する。

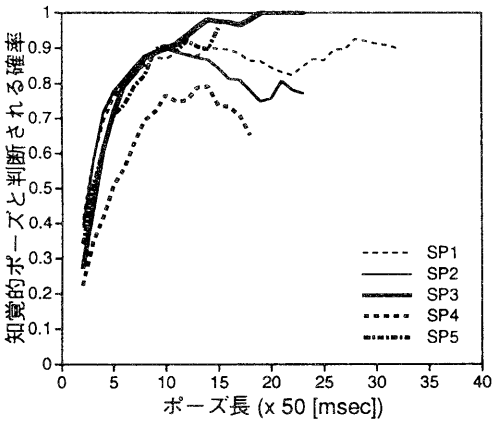

図 2. 音響的ポーズが知覚的ポーズと判断される確率

表 3. 間投詞の種類と話者別頻度

	SP1	SP2	SP3	SP4	SP5
あ [~]	2	1	3	7	16
あの [~]	0	0	28	11	12
う [~]	0	1	0	6	3
う [~] ん	0	3	0	0	0
う [~] んと	0	0	1	0	0
え [~]	24	51	1	35	9
え [~] と	33	1	10	4	0
お [~]	0	2	0	14	11
その [~]	13	1	4	4	0
ま [~]	30	28	88	43	34
ん [~]	0	2	2	3	0
ん [~] と	0	1	1	0	0
合計	102	91	138	127	85
音声長 (秒)	742	570	680	352	519

5.2 つなぎ語

つなぎ語の定義を厳密に下すことは難しいので，ここでは活用語の直後に配置される語，及び文頭に現れる語の幾つかを対象とした。具体的には，**表 4，5** に示す言語表現を対象とした。表中「活」とは活用語を意味する。文頭に来るつなぎ語として，「で [~]」の出現頻度が他と比べて非常に高い。文頭に来ると言うことは，ポーズ前後に発声されやすいことを意味し，知覚的ポーズとの相関が予想される。この点からの考察については**第 6 節**で述べる。

5.3 文末表現 (終助詞)

一方，文末表現としては終助詞とポーズとの相関を見る必要がある。文献 [6] によれば，話し言葉に頻出する文末助詞として「ね [~]，よ，の，か」が上げられている。そこで本研究では，この 4 種類の終助詞に限定してその出現頻度を集計した (**表 6** 参照)。表より，終助詞としては「ね [~]」,「か」が多く使用されていることが分る。特に「ね [~]」は殆んどの場合「~ですね」の形で用いられており，大多数を相手にする講演において，相手に念を押しながら話しを進めている様子が伺われる。これら終助詞と知覚的ポーズとの関係についても**第 6 節**で述べる。

表 4. 活用語直後に発声されたつなぎ語とその頻度

	SP1	SP2	SP3	SP4	SP5
活+て	98	59	102	34	46
活+ても	5	6	10	0	8
活+ので	13	8	13	4	12
活+けれど (も)	9	7	19	10	3
活+けど (も)	7	6	17	4	21
活+が	5	11	3	5	5
活+から	4	0	7	0	7
活+し	2	2	4	0	0
音声長 (秒)	742	570	680	352	519

表 5. 文頭に来るつなぎ語とその頻度

	SP1	SP2	SP3	SP4	SP5
で [~]	18	16	21	4	10
そこで	1	1	2	0	0
それで	0	5	1	0	1
それから	0	0	0	6	0
ですから	4	0	0	0	0
けど	1	0	0	0	0
でも	0	0	1	0	0
音声長 (秒)	742	570	680	352	519

表 6. 着目した終助詞の種類とその頻度

	SP1	SP2	SP3	SP4	SP5
ね [~]	9	1	10	18	47
よ	0	0	1	0	0
の	0	0	0	0	0
か	9	4	13	9	19

5.4 言い直し/言い淀み/繰り返しの類

自然発話の音声では，種々の原因により，言い直し (及びそれに類する現象) が頻出する。ここでは，以下に示す現象に着目して集計を行なった。但し，今回対象とする現象は，単語・句レベルでの言い直しを対象とし，文レベルでの現象は (議論が複雑になるため) 無視することにした。

- **繰り返し**
  同一単語 (句) を連続して発声する。間に間投詞・つなぎ語の挿入を許す (以下同様)。
- **言い直し**
  同一単語 (句) を連続して発声する。但し，語尾変化或は助詞の変化を伴うもの。
- **言い淀み**
  単語 (句) 途中で発声を止め (word fragment)，再度同一単語 (句) を発声する。
- **言い換え**
  一度発声した単語 (句) と同一カテゴリと判断される単語 (句) を発声する。

表 7. 言い直し現象の話者別頻度

		SP1	SP2	SP3	SP4	SP5
繰り返し	±0	8	2	7	2	3
	+α	7	1	5	3	3
言い直し	±0	3	1	6	0	4
	+α	0	0	0	0	0
言い淀み	±0	9	9	5	3	15
	+α	3	0	3	0	1
言い換え	±0	7	1	4	2	3
	+α	0	0	0	0	0
言い間違え	±0	1	0	1	0	0
	+α	0	0	0	0	0
合計		38	14	31	10	29
音声長 (秒)		742	570	680	352	519

- 言い間違え
  明らかに意図したものとは異なる単語 (句) を発声したために, 訂正して正しい単語 (句) を発声する。

以上 5 種類の現象は各々, 更に 2 つに分類される。即ち, 繰り返されて発声された語 (句) に対して付加情報を明示的に与えている場合と, そうでない場合である。付加情報が加わる場合は, 語 (句) の繰り返しの際に当然のことながら, 間投詞・つなぎ語以外の要素が挿入されることになる。**表 7** に分析結果を示す。表中+α, ±0 は各々, 付加情報の有/無を意味している。これらの現象と知覚的ポーズとの相関については**第 6 節**で述べる。

5.5 その他

話し言葉特有の現象としては, その他に "倒置" が上げられる。しかし, 今回使用した音声資料には, 倒置現象 (但し, ここでは述語の後に, その述語に係るべく主語, 目的語が出現するもののみを対象とした) は **SP3** に 3 回現れただけであった。[3] によれば, 対話音声には約 50 文に一度倒置現象が現れると記述されてあるが, 今回の分析結果は, 明らかにそれを下回るものであり, 情報の一方通行である講演音声の特徴として考慮されるべき点である。

6 知覚的ポーズと相関の高い音響的/言語的現象

以上, 知覚的/音響的ポーズ及び種々の言語現象を観察してきたが, 本節では, 知覚的ポーズと相関の高い音響的/言語的現象を検証することにする。

6.1 知覚的ポーズと音響的ポーズ

表 1 で各話者の発話速度を示した。しかしこの数値は, 音声長に対してもモーラ数に対しても何ら後処理を行なっていない。話し手から聞き手に音声を媒体として情報が伝達される過程を考えると, 人間が感じる「間」とは, 情報伝達の区切り, 即ち聞き手にとって情報が入力されていない時間帯であると考えられる。逆に, 情報が入力されている時間帯において, 最終的にそこが間投詞であった場合を考える。間投詞は話し手の発話作業と思考作業との同期をとるために発せられるものであり, 一般的には, 聞き手がメッセージ内容を理解するに当って, 最終的に切り捨てられる部分である。以上の考察より, 以下に示す実効的発話速度を再定義する。

表 8. 各話者別の実効的発話速度 [mora/sec]

SP1	SP2	SP3	SP4	SP5
8.69	8.24	9.11	7.54	9.10

$$\text{実効的発話速度} = \frac{[\text{全モーラ数}] - [\text{間投詞部分のモーラ数}]}{[\text{全音声長}] - [\text{全知覚的ポーズ長}]}$$

これは, 話し手から聞き手に伝えられる単位時間当りの (意味的なレベルでの) 言語的情報をより直接的に反映した数値となると考えられる。なお, 繰り返し/言い直し部に対しては, (結果的に)「強調」などの効果をもたらすと考えられるため, 削除の対象とはしなかった。

表 8 に新たに定義された実効的発話速度を各話者毎に示す。この定義に従えば本研究で使用した音声資料は, **[SP3, SP5]→[SP1, SP2]→[SP4]** の順に情報伝達の速度が低下していることになる。この結果を踏まえて**図 2** を考察すると, 発話速度の高い話者ほど, 音響的ポーズを知覚的ポーズと判断する確率が高くなっていることが分る。特に最も発話速度の低い **SP4** に対しては, 音響的ポーズに対して「間」を感じる確率が著しく低くなっている。これはポーズ節を単位とした音声処理方式を考える場合, 発話速度を考慮に入れる必要があることを示唆する。また今回の分析では発話速度を各話者毎に平均して求めたが, 当然のことながらこれは動的に変化するものである。そのため同一話者に対しても, その発話スタイルに動的に適応した形でポーズを検出・活用する必要があると考えられる。

6.2 知覚的ポーズと種々の言語現象

まず各話者別に, **第 5 節**で考察した「間投詞/つなぎ語/終助詞」が知覚的ポーズの直前あるいは直後に現れた回数を知覚ポーズ節数と共に**表 9** に示す。表より, **SP4** における知覚的ポーズは**第 5 節**で考察したような言語表現を伴うことが, 他と比較して極めて多いことが分る。前節での考察を考慮すると, 実効発話速度が早い場合は, ポーズと言う音響的な要因によって音声の処理単位 (区切り) が形成され, 発話速度が遅い場合は, マーカとなるべき語 (言語的表現) によって音声の区切りの決定が行なわれているものと示唆される。即ち, 音声の区切りを決定する要因として音響的要因と言語的要因があり, 発話速度を一つのパラメータとして優先的に扱われるべき要因が決定される, と言うことである。

次に, 全話者に対して計 20 回以上出現している「間投詞/つなぎ語/終助詞」の全種類に対して, 知覚的ポーズの直前/直後に位置している割合を調べた。結果を**表 10** に

表 9. 知覚ポーズ節数と間投詞/つなぎ語/終助詞

	SP1	SP2	SP3	SP4	SP5
出現回数	153	141	120	128	133
知覚ポーズ節数	299	266	296	142	197

表 10.「間投詞/つなぎ語/終助詞」が知覚的ポーズの直前/直後に位置する割合

	SP1	SP2	SP3	SP4	SP5	率 [%]
あ	2/2	0/1	3/3	3/7	10/16	62.0
あの	0	0	15/28	10/11	5/12	58.8
え	21/24	50/51	0/1	30/35	8/9	90.8
えと	31/33	1/1	9/10	4/4	0	93.8
お	0	1/2	0	10/14	10/11	77.7
その	8/13	1/1	1/4	4/4	0	63.6
ま	19/30	20/28	46/88	15/43	15/34	51.6
率 [%]	79.4	86.9	55.2	64.4	58.5	

	SP1	SP2	SP3	SP4	SP5	率 [%]
+て	22/98	10/59	25/102	5/34	6/46	20.0
+ても	4/5	3/6	3/10	0	4/8	48.3
+ので	6/13	6/8	9/13	2/4	7/12	60.0
+けれど	6/9	5/7	14/19	8/10	3/3	75.0
+けど	5/7	1/6	13/17	2/4	10/21	56.4
+が	2/5	7/11	0/3	3/5	4/5	55.2
率 [%]	32.8	33.0	39.0	35.1	35.7	

	SP1	SP2	SP3	SP4	SP5	率 [%]
で	13/18	16/16	20/21	3/4	9/10	88.4
率 [%]	72.2	100	95.2	75.0	90.0	

	SP1	SP2	SP3	SP4	SP5	率 [%]
ね	4/9	0/1	6/10	13/18	31/47	63.5
か	0/9	1/4	2/13	2/9	1/19	11.1
率 [%]	22.2	20.0	34.8	55.6	48.5	

示す。但し，表中の「間投詞／つなぎ語／終助詞」はそれぞれ省略形で示されており，**N/M** とは全体数 (=**M**) のうち，**N** 個が知覚的ポーズの前後に出現していることを示す。いる。また，話者別に (語間の) 平均をとったものを最下行に，語別に (話者間の) 平均をとったものを再右列に「率 [%]」として示す。表より，間投詞「え [~]」,「え [~] と」やつなぎ語「で」が音声入力を区切り，(大きな) 処理単位を構成するために使用される可能性が非常に高いことが分る。特に「で」は，論旨をある流れに沿って次々と述べるために使われており，講演調の話し言葉の特徴として上げることができると考えられる。

7 まとめ

講演調の話し言葉を音響的/言語的に分析することによって，人間が感じる「区切り」が何によって引き起こされるかを観察した。その結果「区切り」を引き起こす要因は発話速度に依存することが示された。発話速度が高い場合は，音響的要因であるポーズに (より) 依存しつつ音声を区切り，(大きな) 処理単位を形成していると考察される。逆に発話速度が低い場合は，ポーズによる影響は小さくなり，間投詞やつなぎ語と言った言語的表現を優先的に利用して音声を区切っている様子が観測された。特に間投詞「え [~]」,「え [~] と」やつなぎ語「で」が出現した箇所は高い確率で入力音声に対する「区切り」となっていることが示された。

しかし，今回行なった実験では「音声を大きな処理単位に分割する」際に用いられる諸要因について観測したに過ぎない。即ち，ポーズ/間投詞/つなぎ語と言った要因が，更に上位の処理 (音声理解) においてどのような役割を果たしているかについては触れていない。また，基本周波数の挙動やパワーの変化がもたらす影響についても触れていない。**第 3 節**でも述べたが，文献 [9] などにおいて，音声の処理単位はポーズや言語表現の他に，イントネーションが深く関与していると内省的に考察されている。基本周波数やパワーの果たす役割 (入力音声の区切りと音声理解の両レベルにおいて) についても実験的に分析する必要がある。**第 4.3 節**で述べた実験計画の不備の他に，以上のような観点からの実験的考察が今後の課題として残った。

参考文献

[1] 森元，村上，"音声対話における言語現象"，日本音響学会誌，vol.50，No.7，pp588-562 (1994).

[2] 竹沢，田代，森元，"音声言語データベースを用いた自然発話の言語現象の調査"，人工知能学会研究会資料，SIG-SLUD-9403-3 pp.13-20 (1994).

[3] 中川，小林，"自然な音声対話における間投詞・ポーズ・言い直しの出現パターンと音響的性質"，日本音響学会誌，vol.51，No.3，pp202-210 (1995).

[4] 竹沢，田代，森元，"自然発話の言語現象と音声認識用日本語文法"，音声言語情報処理研究会資料，6-5，pp.27-34 (1995).

[5] K.Takagi and S.Itahashi, "Effectiveness of Pause Information in the Content Word Detection of Spoken Dialogues," Proc. ESCA. EUROSPEECH'95, pp.19-22 (1995)

[6] "談話語の実態"，国立国語研究所報告 8 (1955).

[7] "話しことばの文型 (I)–対話資料による研究–"，国立国語研究所報告 18 (1960).

[8] "話しことばの文型 (II)–独話資料による研究–"，国立国語研究所報告 23 (1963).

[9] 伊佐早敦子，"話しことば序"，国語国文，Vol.22，No.3，pp.49-67 (1953).

[10] 宇野義方，"話しことばの文法"，言語生活，No.66，pp.28-35 (1957).

[11] 大石初太郎，"話し言葉とは何か"，話し言葉 (「ことば」シリーズ 12)，文化庁 (1980).

[12] 石井久男，"はなしことばにおける文法"，言語生活，No.406，pp.34-41 (1985).

[13] 三尾砂，"話言葉の文法 (言葉遣篇)"，くろしお出版 (1995，但し，1941 年に出版されたものの復刻版).

[14] "パネル討論：話し言葉の構築は可能か？"，音声言語情報処理研究会資料，6-8，pp.51-60 (1995).

[15] 伊藤，秋葉，上篠，田中，"休止を区切りとした対話処理"音声言語情報処理研究会資料，7-21，pp.135-138 (1995)

[16] 保坂，衛藤，"話しことばにおけるポーズ節の考察" 情報処理学会全国大会論文集，2Q-6，pp.110-111 (1994)