

仮想残響場での話者原音声の分析  
 ～コミュニケーションの意志が音声の時間構造に及ぼす影響～\*

◎ 藤井 無限 福留 公利 (九州芸術工科大学)

1. はじめに

話者周辺の音環境が変化した場合に、話者は、発話音声を変化させ、音声コミュニケーションを行っている。このような、聴覚-発声のフィードバックによる発話の変化に関しては、Lombard 効果が有名である。これは、騒音中で発話した場合に音声に変化する現象で、古くから、さまざまな研究がなされている。近年では、Junqua らによる分析が大規模で定量的である [1]。また、更に Junqua らは、被験者にノイズを聞かせながら、リスト読み上げデータベースを作成するだけでなく、コミュニケーションの効果を考慮した Lombard 音声のデータベースを作成し分析を行っている [2]。

本研究では、音環境として残響音場に着眼し、Junqua 同様コミュニケーションに注目する。すなわち、コミュニケーションの意志を反映した音声のデータベースを作成し、音声の時間構造に対する変化を分析する。また、残響のある環境での、コミュニケーションの意志を、「離れた場所にいる相手に聞き取り易いように話す努力である」と定義する。

残響音場で発話された話者の原音声进行分析するために、2つの無響室と実時間たまたみ込み装置を用いて、仮想的に話者に残響を提示し、話者原音声を収録する方法を考えた。また、収録後の分析は、日本語の拍と母音エネルギー重心点間時間長 ( $D_{cegv}$ ) に関して行った。これは、文献 [1] で音素の長さが非線形に変化することが報告されており、拍に関してはどうか? という点に関心があるからである。また、分析のために、2種類のラベリング方法を用いた。

2. 話者原音声収録実験

2.1 実験システム

図 1は、残響音場で発話された話者原音声を収録するための対話型システムの概略図である。話者原音声を収録するために、無響室を用いており、聞き手と話者を想定した対話型のシステムになっている。この実験システムにより、仮想的に話者と聞き手を、同一の残響音場に置き、対話形式で実験を行うことができる。なお、想定している話者と聞き手の距離は 5m である。大無響室と小無響室の音声は、まず原音声として DAT に録音され、その後音声にインパルス応答がたまたみ込まれ小無響室の聞き手、大

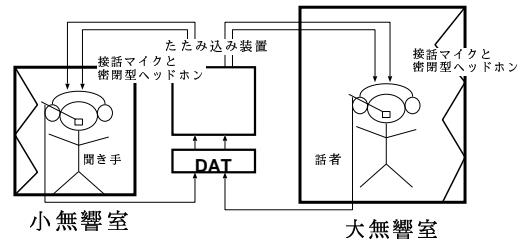


図 1. 対話型実験概略図

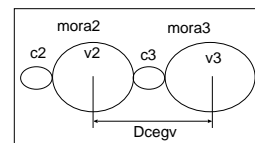


図 2. 分析対象とするモーラ

無響室の話者の密閉型ヘッドホン内にフィードバックされる。

本研究では、以下の 4つの音響伝達経路を模擬したものを被験者に提示する。1) 残響室での話者の口から外耳道の入口までの音響伝達経路。2) 残響室での相手話者 (聞き手) の口から話者の外耳道入口までの音響伝達経路。3) 無響室での話者の口から外耳道の入口までの音響伝達経路。4) 無響室での相手話者 (聞き手) の口から話者の外耳道入口までの音響伝達経路。このそれぞれを模擬するためのインパルス応答を、実際の部屋で測定した。なお、残響室の残響時間は、5m の距離で測定したインパルス応答から算出してエネルギーが 60dB 減衰するまでの時間を測定して 1.6 秒であった。

2.2 実験方法

実験に使用した語彙リストは、母音部エネルギー重心点間時間長 ( $D_{cegv}$ ) の分析を中心に行うため、加藤、橋本と同様なものを作成した [3]。加藤らは、母音部エネルギー重心点間時間長に関してリスト読み上げ音声で定量的な分析を行っている。この語彙リストは「それは、こ〇〇めんかいです。」という下線部無意味 7モーラを含む文の〇〇の部分について変化させたものである。1) 先行母音  $v_2$  が違う場合、2) 後続母音  $v_3$  が違う場合、3) CEGV 間の子音  $c_3$  が存在せずに撥音あるいは母音のみで続く場合、4) 先行の子音  $c_2$  が違う場合、5) CEGV 間の子音  $c_3$  が違う場合、6) CEGV 間の子音  $c_3$  が促音になった場合、7) 後続の母音  $v_3$  が無声化した場合。

\* Analysis of an original speech uttered in a simulated reverberant sound field  
 -The influence of the communication effort upon the time structure of the original speech-  
 By MUGEN Fujii, KIMITOSHI Fukudome (Kyushu Institute of Design)

ここで、 $c_2, c_3, v_2, v_3$ は図2に記した通りである。

実験は、話者1名、聞き手2名を用いて行った。さらに、話者に関しては比較のため、あらかじめ今回の「それは、こ〇〇めんかいです。」をリスト読み上げで5セット収録した。1セットは、残響を提示しながら発話した場合と残響のない状態で発話した場合の2試行を含んでいる。また、この内1セットは、話速に関して有意な違いが出ていたので無効とした。

本実験は、話者と聞き手の対話形式で行う。作成した疑似会話文を以下に記す。ここで、聞き手はランダムに数字が配列された数表を持っており、話者は語彙リストに番号が打ってある表を持っている。

lis ×番目の文章は何ですか？

sp ×番ですか？

lis はい、×番です。or いいえ、×番です。

sp ×番は、[ポーズ]それは、こ〇〇めんかいです。  
(1、分析対象音声)

lis それは、こ〇〇めんかいですか？

sp はい、[ポーズ]それは、こ〇〇めんかいです。or  
いいえ、[ポーズ]それは、こ〇〇めんかいです。  
(2、分析対象音声)

この疑似対話文に従い、聞き手が話者に尋ね、話者が答えるという会話を行う。対話型実験1セットは残響のある場合と、残響のない場合の2通りの試行で構成される。実験は2セット行い各セットごとに聞き手を交換した。なお、聞き手には、聞き取れた単語を書きとってもらっている。こうすることで、聴き難い環境で、相手に分からない情報を伝えるという状態を疑似的に再現する。

この実験の結果、条件1「残響がない条件で発話された、リスト読み上げ音声(以降、al)」、条件2「残響がある条件で発話されたリスト読み上げ音声(rl)」、条件3「残響がない場合での対話音声(ac)」、条件4「残響がある場合での対話音声(rc)」の4条件に対しそれぞれ、4回のセンテンスが収録された。以降この音声に対し、それぞれ分析を行う。

### 3. 話者原音声の分析

#### 話速に関する分析

各条件の音声に対して話速を分析し平均処理した結果は、alが6.78(moras/sec)、rlが6.63(moras/sec)、acが6.33(moras/sec)、rcが5.34(moras/sec)であった。残響音場に置けるコミュニケーションの意志の効果でrcの話速が極端に遅くなっていることが分かる。なお、残響音場での対話実験中普通の話速では、コミュニケーション不可能であることが内感報告として分かっている。

また、単語全体の平均エネルギーに関しては、rcが一番高く、alが一番小さかった。残響のある聞きにくい環境のなかで、遠くの聞き手に伝えようと、エネルギーの高い発話をしていることが分かる。言い替えば、全体的に話速が遅くなり、発話のエネルギーが高い条件で音声の時間構造がどのように変化するかを分析する。

#### ラベリング方法について

前節の実験によって得られた原音声データに対し、手動でラベリングを行い、分析を行う。ラベリングの方法として、2通りの方法を用い分析を行う。1つは、母音の遷移の始まる点に注目したラベリング手法である。図3のホルマント周波数の遷移とスペクトル構造から母音の遷移の始まる点を探し、その点を子音の始まりとしている。この方法により、子音に調音運動の始まりの時間が分かると考えている。図3の■の点が、上述した点で■間の距離がこのラベリング手法でのモーラ長となる。

もう1つは、母音のパワーが最小になる点を目安に、子音との境界を決定したもので  $D_{cegv}$  の算出を目標としている。図3の口がこのラベリング手法での境界線である。

#### モーラ長の分析

「こ〇〇めんかいた」の〇〇部分2モーラに関して、パワに注目したラベリング手法それぞれに対して、モーラ長を分析した。結果の一例として、1)先行母音が違う場合の結果に対して、対象2モーラのモーラ長の差を、表1、表2に記す。これは、1セットの結果を記したものである。

分析の結果について述べる。

パワに注目したラベリングに基づく分析の結果以下のようなことが分かった。1)〇〇部分の音素により、モーラ長の単語長に対する割合が変わり、音素が変わってもモーラ長が等時的であるとは、言い難いこと、2)2モーラ目と3モーラ目の長さが等時に

表1. パワに注目したラベリングに基づく分析結果: 「こ〇〇めんかいた」の第2モーラ長-第3モーラ長

	al(ms)	rl(ms)	ad(ms)	rd(ms)
baba	21	35	-14	-11
biba	-7	6	-16	-25
buba	-16	-27	4	20
beba	1	-5	-11	-5
boba	44	-6	-17	59

表2. 遷移に注目したラベリングに基づく分析結果: 「こ〇〇めんかいた」の第2モーラ長-第3モーラ長

	al(ms)	rl(ms)	ad(ms)	rd(ms)
baba	-8	-3	10	1
biba	-3	-2	6	2
buba	4	4	-1	1
beba	-2	1	-9	1
boba	8	-9	12	2

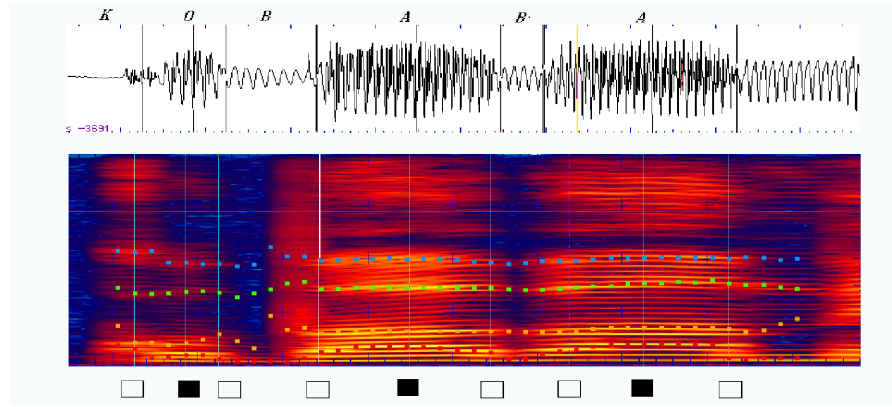


図 3. 「こ〇〇めんかい」のラベリングの例, 「こばばめんかい」に対しラベリングをしているところ。

なるとは言い難いこと、3) 2 モーラ目と 3 モーラ目の 2 拍に負の相関がみられ、2 拍で長さを一定にしようとする傾向があること。4) モーラ長に対してコミュニケーションの意志が、明らかに影響するとは、このデータだけでは断定出来ないこと。

また、母音遷移に着目したラベリングに基づく分析の結果以下のことが分かった。1) 〇〇の部分の拗音である場合以外は、〇〇の 2 拍に等時的な傾向がみられること、2)、1) の等時傾向は、コミュニケーションの意志による話速変化、エネルギー変化の影響をほとんど受けないこと。

#### 母音エネルギー重心点間時間長

残響音場で、全体的に、話速が遅く、エネルギーの高い発話をするので  $D_{cegv}$  にどのような影響があるのかを調査する。図 4 は先行母音を代えた場合の  $D_{cegv}$  の単語長に対する割合をプロットしたもので、図 6 は先行子音を代えた場合、図 7 は間の子音を代えた場合の  $D_{cegv}$  の単語長に対する割合をプロットしたものである。分析結果から 1) 話し方が変化することによる  $D_{cegv}$  への影響は、単語長に対する割合の差で、ほぼ 2% 以内であること、2) 母音重心間の子音を変化させる場合が、他の場合より  $D_{cegv}$  に対する影響が遥かに大きいこと<sup>[3]</sup>、3) 話し方が変化する影響よりも遥かに、母音重心間の子音の影響のほうが大きい (ほぼ 6%) ことが分かる。

## 4. 議論

前節の結果に対して議論する。

図 8 は、発話のリズムに関する comb モデルを示したものである<sup>[4][5]</sup>。  $T_i$  は言語運動中枢において次ぎの指令が出されるまでの時間、  $P_i$  は調音現象の開始時刻、  $X_i$  は言語運動中枢から末梢である音声器官への運動指令が伝達し調音が実現するまでの時間である。等時性といった場合、  $T_i - T_{i-1}$  の間隔が等しいことを意味し、  $X_i$  による違いがほとんどないと仮定すれば、  $P_i - P_{i-1}$  にも等時的な傾向がみられることになる。このモデルを仮定して以下議論を勧

める。

母音遷移に注目したラベリングを行うことで、  $P_i$  の時間に、ほぼ等しい点を求めることが出来る。分析の結果、この点  $P_i$  の間隔は 10ms 以内であった。また、残響音場でのコミュニケーションの意志の影響で単語全体の話速が遅くなり、  $T_i - T_{i-1}$  の間隔が大きくなるが、  $P_i - P_{i-1}$  の等時傾向には、影響がないと考えられる。

パワーに注目したラベリングによる分析の結果と母音遷移に注目したラベリングの結果を比べると、前者のみ 1 モーラ単位の等時性が成り立たないことから、調音運動の開始点  $P_i$  からパワーが最小になるまでの時間は、音韻環境によって異なるといえる。また、断定は出来ないが、残響音場でのコミュニケーションの意志がこの時間長には、ある程度、影響するのではないかと考えられる。

上述の議論より、1 つの仮説として次ぎのようなものが考えられる

[仮説] 母音のホルマント、及びスペクトル構造が変化し始める点を聴き、この点でタイミングを取り、言語中枢からの指令が出される。このプロセスには、残響音場でのコミュニケーションの意志は全体の話速を変化させる以外関与しない。

## 課題

本研究は、無意味単語の特定モーラ部分に対してのみ分析を行い議論をした。この点に関しては、有意義なセンテンスを用いたより自然な会話文で、原音声収音実験を行うことが必要である。また、その収音データに対しアクセント型、ポーズ、統語構造のさまざまな観点から分析し議論をすることが必要である。

被験者、セット回数がまだ十分でないことも問題で、データ数をより充実させ分析を試みる必要がある。

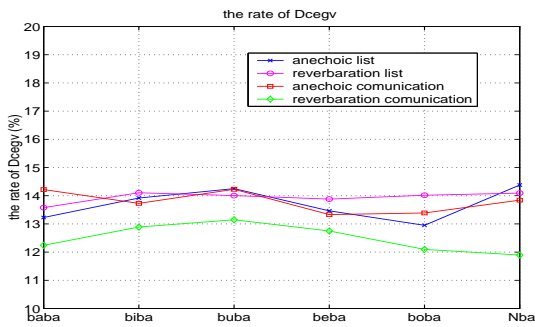


図 4. (1) 先行母音が違う場合の  $D_{cegv}$  の単語長に対する割合 (「こ〇ばめんかい」の〇の母音が違う場合)

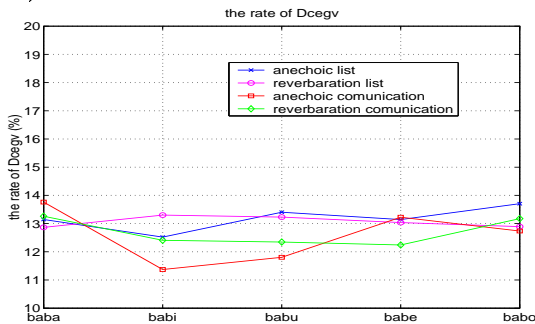


図 5. (2) 後行子音が違う場合の  $D_{cegv}$  の単語長に対する割合 (「こば〇めんかい」の〇の母音が違う場合)

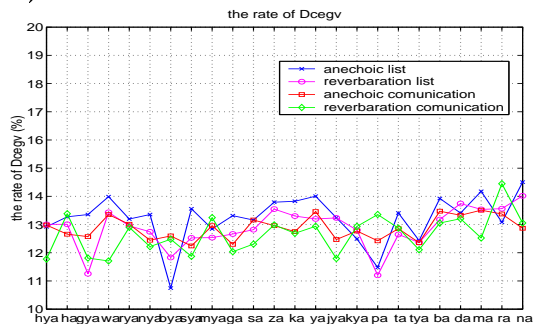


図 6. (4) 先行子音が違う場合の  $D_{cegv}$  の単語長に対する割合 (「こ〇ばめんかい」の〇の子音が違う場合)

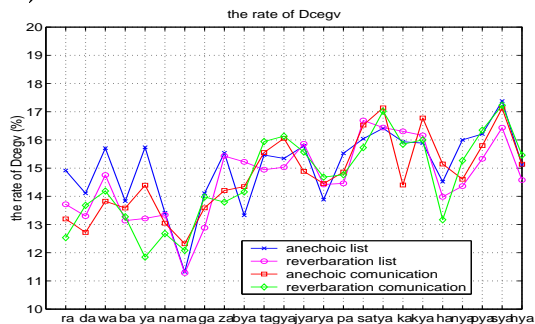


図 7. (5) 間の子音が違う場合の  $D_{cegv}$  の単語長に対する割合 (「こば〇めんかい」の〇の子音が違う場合)

## 5. まとめ

本研究では、実時間たたみ込み装置を用いて、仮想残響場での対話型実験を行った。そして、パワに

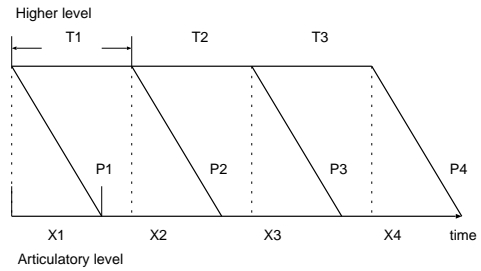


図 8. 発話リズムに関する comb モデル (文献 [4][5] より)

注目したラベリング手法を用いた分析の結果以下のことが分かった。(1) 無意味単語「こ〇〇めんかい」の〇〇部分 2 モーラどうしに音素を変化させた場合、等時傾向がみられないこと、(2) 残響音場でのコミュニケーションの意志は、全体の話速に対しては強く影響し、 $D_{cegv}$  へ及ぼす影響は、単語長からの割合の差で、ほぼ 2% 以下であること、(3) 母音間の子音の違いがもたらす影響はほぼ 6% であること

また、母音遷移に注目したラベリングを用いた分析の結果、無意味単語「こ〇〇めんかいです。」の〇〇部分の隣接 2 モーラどうしが、等時になる傾向がみられ、音素環境を代えてもモーラ長が等しい傾向があった。そして、この傾向はコミュニケーションの意志の影響を全く受けないことが分かった。

## 参考文献

- [1] Jean-Claude Junqua, "The Lombard reflex and its role on human listeners and automatic speech recognizers", J. Acoust. Soc. Am. **93**(1) pp.510-524 (1993).
- [2] Jean-Claude Junqua, Steven Fincke and Ken Field, "INFLUENCE OF THE SPEAKING STYLE AND THE NOISE SPECTRAL TILT ON THE LOMBARD REFLEX AND AUTOMATIC SPEECH RECOGNITION", Proc. ICSLP vol2, pp467-470(1998)
- [3] 加藤雅代、橋本新一郎, "母音重心点エネルギーに着目した日本語のリズム規則について", SP92-11, pp33-40(1992)
- [4] 渡辺義尚、粕谷英樹, "生成過程を考慮した朗読及び対話音声の時間構造の検討" SP94-103, pp37-44(1995)
- [5] 川崎春子, "音声の時間制御に関するモデルと実測データ", 音響学会誌 39, 389-396(1983)