



# Perceptual Characteristics of Temporal Structures in Speech : Towards Objective Assessment of Synthesis Rules

加藤, 宏明

---

(Degree)

博士 (工学)

(Date of Degree)

1999-03-31

(Date of Publication)

2009-10-14

(Resource Type)

doctoral thesis

(Report Number)

甲1927

(JaLCD0I)

<https://doi.org/10.11501/3156328>

(URL)

<https://hdl.handle.net/20.500.14094/D1001927>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



氏名・（本籍）	かとうひろあき 加藤宏明	（静岡県）
博士の専攻分野の名称	博士（工学）	
学位記番号	博い第149号	
学位授与の要件	学位規則第4条第1項該当	
学位授与の日付	平成11年3月31日	
学位論文題目	Perceptual Characteristics of Temporal Structures inSpeech : Towards Objective Assessment of Synthesis rules (合成規制評価のための音声の時間構造知覚の研究)	
審査委員	主査 教授 匂坂芳典	
	教授 北村新三	教授 安藤四一

### 論文内容の要旨

本論文は、自然な音声の合成への指針を得るために行なわれた音声の時間構造知覚の研究をまとめたものである。任意のテキストから音声を生産する規制音声合成では、自然性の高い音声を実現するために、時間構造を自然発話の場合に近いものに制御する方法がとられている。この制御に要求される精度は、合成された音声の最終的な受益者である人間が許容する誤差の量によって決まる。したがって、最適な制御を行うためには、種々の誤差に対する人間の許容特性の知識が不可欠であるが、これまでは極めて限られた例の実験報告が得られるのみである。

本研究では、人間が音声の時間構造の乱れを感じ取る際に影響される所要因を心理実験により系統的かつ多面的に調査することで、合成規制における時間構造制御に必要な知識の集積を図り、さらに、それらの知識を合成規則の評価に反映させるための枠組みを提案し、実際の評価モデルを試作することで得られた知識の有用性を実証した。また、実験で得られた知識が今後分野を越えて広く使われることを想定し、心理音響学的な解釈の可能性を検討することによって当該知識の適用可能領域についての指針を与えた。本論文は、以下に述べる8章からなる。論文の構成を図0.1に示す。

第1章では、まず音声コミュニケーションにおける時間構造の役割とそれに対する知覚の特性を概説し、次に従来の音声合成規制の評価方法を概観し、それらが内包する知覚的側面からの不十分さを明らかにすることによって、本研究で提起する問題の所在を示す。

第2章から第5章では、第1章で明らかにされた音声聴取における時間知覚に関する問題に取り組むため、“変形に対する感じやすさ”という切り口により種々の条件の音声を実験刺激として知覚特性を調査する。章が進むにしたがい、音声刺激条件は基本的なものから複雑なものへと移行する。

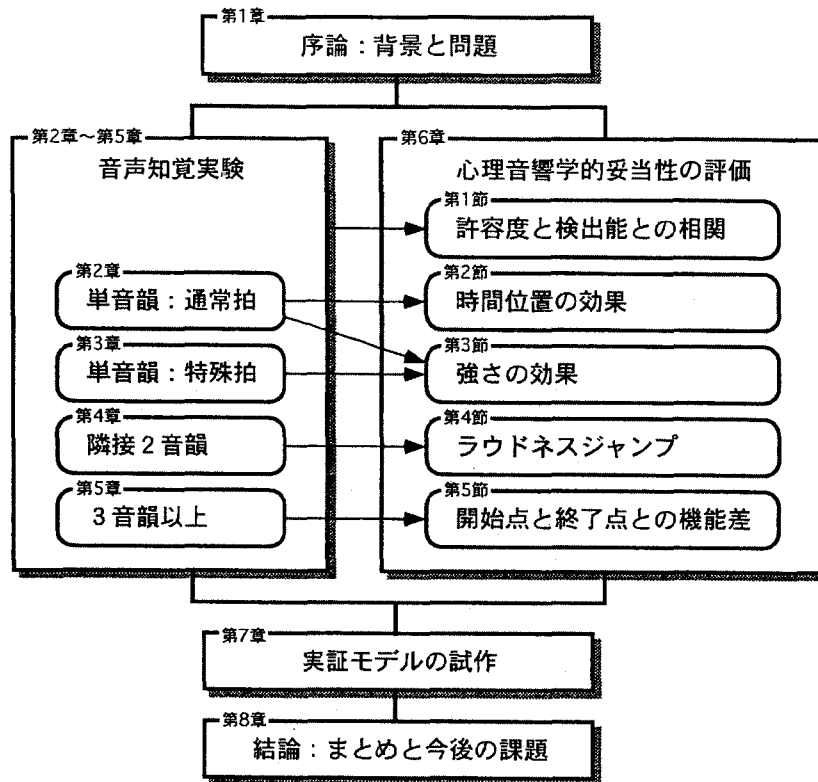
第6章では、第2章から第5章の音声知覚実験で得られた知見の心理音響学的妥当性を音声の時間構造を模した非音声をを用いた実験により検証する。この章での調査は、音声実験において得られた知識の適用可能領域を見積もるために重要と考えられる。

第7章では、ここまでの実験結果をまとめ、人間の知覚特性を考慮した音韻長歪みの評価モデルを

試作することによって、得られた知見の合成規則の評価における有用性を実証する。

第8章で以上の結果を統括し、今後の課題と将来の方向を展望する。

本研究の目的は、言い換えれば、人間が聴覚を通じて外界の変化を感じ取る能力の鋭敏さと遅鈍さが音声の知覚にいかにか巧みに組み込まれているかを知ることでもある。今後ますます需要が高まっていくであろう人と機械との音声を媒介としたコミュニケーションの場面において、このような人間の知覚メカニズムの知識が使用者に負担を強くない技術を確認するための基盤となることが期待される。以下に章毎の題を示し内容を概説する。



## 第1章 序論

第1章では、まず研究の動機と目的を述べた。音声コミュニケーションにおいて、時間構造は情報交換を円滑に行うために送り手と受け手とが共有すべきプロトコルの役割を果たす。しかし、機械が送り手となる音声合成では、時間構造は受け手の人間にとって常に最適に制御されるわけではない。すなわち、物理的時間と人が感じる時間すなわち心理的時間とは必ずしも線形に対応するわけではないため、現在広く行われている物理的時間のみに頼る客観評価法では真に最適な制御へのガイドラインを与えることはできない。本研究は音声の時間的歪みに対する人の知覚特性を系統的に調査することで、人の感じ方を取り入れた客観評価法の確立を図るものである。

本章では、加えて、この論文で用いた心理実験の枠組みを概説し、最後に、論文の構成を述べた。

## 第2章 単母音長変形に対する許容特性

—単一音韻長変形の知覚：通常拍—

第2章から第5章では音声の時間構造に対する知覚感度の測定を行なった。

第2章では、最も基本的な実験条件として、日本語の通常拍、すなわち子音と母音とが規則的に交替する音声に含まれる単一の音韻に歪みが加わった場合を対象とした。その結果、母音長の変化に対す

る知覚の感度に影響を与える要因として、変化させた母音の種類（例：‘a’か‘i’か）、その単語内での時間的位置（例：1文字目か3文字目か）、後に続く音韻の種類（例：有声音か無声音か）、が挙げられることが分かった。変化させた母音の元々の長さの違いは知覚感度に影響を与えなかった。

### 第3章 特殊拍を含む区間長の変形に対する許容特性

#### －単一音韻長変形の知覚：特殊拍－

日本語には通常拍に加えていわゆる特殊拍が存在し、その出現頻度も高い。第3章では、第2章と同様の方法で、特殊拍、すなわち撥音、促音、長音、無声化母音を含む音韻の変形に対する知覚感度を測定した。その結果、特殊拍では許容される時間歪みの範囲が通常拍の場合よりも一般に広く、それは特殊拍の定常部分の心理音響的性質、すなわち小さなラウドネス値、と特殊拍に固有の時間構造に起因するものとして説明できることが分かった。

### 第4章 隣接2音韻の時間変形に対する許容特性

#### －隣接2音韻長変形の知覚－

合成音における時間歪みは一般には単一音韻のみではなく複数音韻に同時に生じるものである。第4章では、複数音韻に生じた歪みの知覚に対する相互作用として、隣接2音韻間での知覚的音韻長補償効果を調査した。その結果、補償効果の生じやすさが、関係する2音韻間の心理音響的特徴の差（ラウドネスの落差）に依存しており、仮名一文字に相当するような音韻論上の単位とは無関係であることが分かった。

### 第5章 音声の時間構造知覚における母音開始点と母音終了点との機能差

#### －多数音韻長変形の知覚－

音声合成における音韻長の制御は、各音韻の開始点と終了点の時間位置を決定することに他ならない。日本語では、子音の終了／開始点は大概母音の開始／終了点に一致する。第5章では、さらに多くの音韻に同時に変形が加わった場合を想定して、母音開始点と終了点との時間構造知覚の手がかりとしての機能について調査した。その結果、音韻長の知覚的測定には母音終了点は開始点同様に重要な役割を果たすが、発話テンポの把握などより広い範囲の処理には母音開始点のみが貢献することが分かった。

### 第6章 音声知覚実験の心理音響学的解釈

第2章から第5章で調査した知覚特性に基づき、次章では所与の音韻長歪みに対する主観値を予測するモデルの構築を目指す。モデルの構築に際しては、より広い適用可能性を確保するため、人間の聴知覚の一般的特性に準拠したものを狙う。このため、本章では、第2章から第5章の各々で観察された音声の知覚特性が、人間の聴覚一般に備わった特性が反映されたものとして解釈可能かどうかを非音声をを用いた心理音響実験により検討した。ここで行なった実証的検討は、音声知覚実験で得られた結果の普遍性の吟味に役立つだけでなく、広く一般的な人間の時間知覚特性の解明にも貢献することが期待される。

### 第7章 音韻聴設定規則の客観評価モデル

第7章では、ここまでの実験結果をまとめ、得られた知見が合成規則の評価に有用であることを音韻長歪みの主観評価モデルを試作することにより実証した。このモデルは、客観的指標のみによりながら、人間の主観に近い音韻長設定誤差の評価を可能にするものである。まず、音声の時間構造表現の一つの枠組みとして、時間－ラウドネスマーカ表現を提案する。これは、ラウドネスの時間変化を音韻毎に量子化したものである。さらに、この枠組みに則って音韻長設定誤差に対する客観評価モデルを実現し、このモデルにより心理実験から得られた主観評価データの予測を行った結果、従来型の

音韻長誤差のみに基づく予測よりも一貫して精度の良い、すなわち人間の評価に近い、成績が達成された。

## 第8章 結論

まず論文を手短に総括し、最後に将来への課題を示して結びとした。

### 論文審査の結果の要旨

本論文は、自然な音声の合成への指針を得るために行なわれた音声の時間構造知覚の研究をまとめたものである。任意のテキストから音声を生成する規則音声合成では、自然性の高い音声を実現するために、時間構造を自然発話の場合に近いものに制御する方法がとられている。この制御に要求される精度は、合成音声の最終的な受益者である人間が許容する誤差の量によって決まる。したがって、最適な制御を行うためには種々の誤差に対する人間の許容特性の知識が欠かせないが、これまでは極めて限られた数の実験報告が得られるのみである。本研究は、音声の時間的歪みに対する人の知覚特性を系統的かつ多面的に調査することで合成規則における時間構造制御に必要な知識の集積を図ったものであり、さらに、それらの知識を合成規則の評価に反映させるための枠組みを提案、実際に評価モデルを試作することで得られた知識の有用性を実証したものである。

第1章では、まず音声コミュニケーションにおける時間構造の役割とそれに対する知覚の特性を概説し、次に従来音声合成規則の評価方法を概観し、それらが内包する知覚的側面からの不十分さを明らかにすることによって、本研究で提起する問題の所在を示している。

第2章から第5章では、第1章で明らかにされた問題に取り組むため、“変形に対する感じやすさ”という切り口により種々の条件の音声を実験刺激として知覚特性を調査している。章が進むにしたがい、音声刺激条件は基本的なものから複雑なものへと移行する。第2章では、まず最も基本的な実験条件として、日本語の通常拍、すなわち子音と母音とが規則的に交替する音声に含まれる単一の音韻の変形について調査し、母音長の変化に対する知覚の感度に影響を与える要因として、変化させた母音の種類（例：“a”か“i”か）、その単語内での時間的位置（例：1文字目か3文字目か）、後に続く音韻の種類（例：有音声か無音声か）、を挙げている。さらに、変化させた母音の元々の長さの違いが知覚感度に影響を与えないことも指摘している。

第3章では、いわゆる特殊拍、すなわち撥音、促音、長音、無声化母音を含む音声区分の変形に対する知覚感度を調査し、許容される時間歪みの範囲が通常拍の場合よりも一般に広いことを示し、その効果について特殊拍の定常部分での音声学的特徴と特殊拍に固有の時間構造の両面から考察し、解釈を与えている。

第4章では、より複雑な条件として隣接する2音韻に生じた時間誤差の知覚に対する相互作用である知覚的音韻長補償効果を調査し、補償効果の生じやすさが、関係する2音韻間の音響的特徴の差（ラウドネスの落差）に依存しており、仮名一文字に相当するような音韻論上の単位とは無関係であることを明らかにしている。

第5章では、さらに多くの音韻に同時に変形が加わった場合を想定し、母音開始点と終了点との時間構造知覚の手がかりとしての機能について調査した結果、音韻長の知覚的測定には母音終了点は開始点と同様に重要な役割を果たすが、発話テンポの把握などより広い範囲の処理には母音開始点のみが貢献することを明らかにしている。

第6章では、第2章から第5章の音声知覚実験で得られた知見の心理音響学的妥当性を音声の時間

構造を模した非音声をを用いた実験により検証している。この章での調査は、音声実験において得られた知識の適用可能領域を見積もるために重要と考えられる。

第7章では、ここまでの実験結果をまとめ、人間の知覚特性を考慮した音韻長歪みの評価モデルを試作することによって、得られた知見の合成規則の評価における有用性を実証している。

第8章で以上の結果を総括し、今後の課題と将来の方向を展望している。

本研究で得られた実験的成果はいずれもこの分野における新知見であり、合成音声の制御のみならず、広くマンーマシンシステムでの音声インタフェースの設計に貢献することが期待される。

このように、本研究は、これまでその必要性は認識されていながら調査が十分に行われていなかった音声の時間構造知覚に関わる新知識を系統的に示し、その有用性を明らかにし、人間の知覚機構の解明および音声システムの設計・評価に関して重要な知見を得たものとして、価値ある集積であると認める。

よって、学位申請者 加藤宏明は、博士（工学）の学位を得る資格があると認める。  
い。