

Title	眼球運動に基づいた発話計画メカニズムの検討
Author(s)	黄, 金峰
Citation	
Issue Date	2018-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/15178">http://hdl.handle.net/10119/15178</a>
Rights	
Description	Supervisor: 党 建武, 先端科学技術研究科, 修士(情報科学)

修 士 論 文

眼球運動に基づいた発話計画メカニズムの検討

北陸先端科学技術大学院大学  
先端科学技術研究科情報科学専攻

黄 金 峰

2018 年 2 月

# 修士論文

## 眼球運動に基づいた発話計画メカニズムの検討

指導教員 党 建武 教授

審査委員主査 赤木 正人 教授  
審査委員 鵜木 祐史 教授  
審査委員 小谷 一孔 教授

北陸先端科学技術大学院大学  
先端科学技術研究科情報科学専攻

1610067 黄金峰

提出年月: 2018年2月

## 概要

目の動きは、読解中の単語認識および発話計画プロセスの脳活動を反映することができる。これまでの研究は、ほとんど孤立的な単語だけを用いて発話計画の潜時時間を調査した。しかし、実環境の発話計画の仕組みを探索することは困難である。本論文では、連続音声を用いて、中国語のテキストを読み上げる際の発話と眼球運動を一致させることで、発話計画のメカニズムを調べた。眼球運動の注視点を用いて読み上げの計画単位を推定し、読み上げ文の各単位について発話計画の潜時時間を測定した。発話の計画単位を推定したところ、大部分の計画単位は文法的な単語であり、その範囲は1音節から3音節までの単語であることがわかった。計画単位の潜時時間は、読みの進行とともに時間軸に沿って徐々に減少する。また、文が2つの部分文から構成されている場合には、潜時時間の減少傾向は、2つの部分文の境界でリセットされたようである。読みの進行に伴う潜時時間の短縮という現象に基づいて、恐らく潜時時間が語彙理解とモーター・コマンド設計という2つの部分からできたと考えられる。この推測は、意味のない文を用いた実験によって確認された。連続音声の場合に、潜時時間は前後単語の語彙理解の影響が顕著であることを明らかにした。発話計画潜時時間の脳活動については、脳電図を導入して実験が行った。その結果、先頭単語を計画する時の脳機能活動図から見ると、Wernickeらが提唱された「言語の脳モデル」をサポートすることができた。また、脳機能活動図に基づき、テキスト朗読の脳ネットが構築された。潜時時間において、複数の脳領野が活発してお互いに情報を交換していることがわかった。

# 目次

<b>第1章</b>	<b>序論</b>	<b>1</b>
1.1	研究背景と目的	1
1.2	本論文の構成	2
<b>第2章</b>	<b>実験設計と分析方法</b>	<b>3</b>
2.1	実験の全体像	3
2.2	眼球運動と音声測定システム	4
2.2.1	眼球運動の校正	4
2.2.2	刺激呈示の流れ	5
2.3	潜時時間の測定方法	5
2.3.1	利用された眼球運動指標	6
2.3.2	各単語の音声切分	7
2.4	計画単位の統計方法	9
<b>第3章</b>	<b>データ分析と結果</b>	<b>10</b>
3.1	実験一：単純文の分析	10
3.1.1	単純文の構築	10
3.1.2	潜時時間の変化傾向	11
3.1.3	計画単位の統計	12
3.2	実験二：複雑文の分析	13
3.2.1	複雑文の構築	13
3.2.2	潜時時間の変化傾向	14
3.2.3	計画単位の統計	16
3.3	実験三：無意味文の考察	17
3.3.1	無意味文の構築	18
3.3.2	潜時時間傾向の変化	18
<b>第4章</b>	<b>潜時時間における脳活動</b>	<b>22</b>
4.1	実験の設計と流れ	22
4.1.1	実験コーパスの構築	22
4.1.2	実験の設計と刺激呈示	22
4.2	先頭単語の分析と結果	24

第5章	全体の考察	26
第6章	結論	27
6.1	本研究で明らかにされたことの要約 . . . . .	27
6.2	今後の課題 . . . . .	27

# 第1章 序論

## 1.1 研究背景と目的

近年、人間の音声生成過程の研究には、生理・物理的な測定だけではなく、脳レベルに対して関心がますます広がっている。そのため、人間が発話音声の形成に至るまで、脳内において行った発話計画のメカニズムを究明することは非常に重要な課題になってきた。そこで、脳電図技術が成熟している現在、脳レベルの発話計画メカニズムを解明することは可能になり、本研究はそれを眼球運動と結合して試みた。

発話計画は一体何のもののでしょう。おおざっぱに言うと、発話計画は人間の音声生成過程において最も重要な一環であり、発話意図から発話運動まで音声出力までの間に位置すると考えられる。読み上げ音声の場合、話者が目で注視した単語を脳の視覚野を通し、角回で音とマッチングしてから、言語運動野で発話運動を計画し、発話・発声器官を制御しながら音声を生成するというステップからなる [9]。このような振る舞いは、脳の中で一体どのように発話を計画されているのであろうか。

これまで発話運動に伴うテキスト読みの音声生成過程を脳電図による分析は困難であったため、主に眼球運動指標に着目して、人間の発話計画メカニズムを推測してきた [2][3][5][6]。言語心理学の研究分野において、文を利用して人間が自然的な読み過程を検討した結果、黙読および音読の時、眼球の動きは最後の音声生成に深い関係があり、人間の発話計画過程を反映できると証明された [5]。しかしながら、これらの研究は朗読する時の産出音声（朗読音声）を考えず、眼球運動指標のみを考察して心理的な発話過程を推測するだけである。朗読音声を考えられていないため、発話計画メカニズムの探究に対して十分ではない。

そのため、音声生成機構の研究において、眼球運動と音声測定を接合し、文字の注視から発話の開始までの反応時間（潜時時間）を測定しながら発話計画過程を調査された。その結果、発話計画の潜時時間に対して様々な影響があることが報告されている。Sternbergらは、遅延命名課題と呼ばれる手続きを使って、英語の単語を発話刺激として用いた実験を行った。その結果、孤立的な単語を用いた場合について、発話計画の潜時時間は単語の長さに比例となり、音節長から影響されているとわかった [3]。また、頻度効果を利用する実験から検討が行われており、その単語の頻度と反比例となることが明らかにした [10]。

これらの研究によると、眼球運動と音声測定を結合してから、発話計画のメカニズムを推定することが可能であると示している。しかし、それらの研究は単語のみを使って発話計画のメカニズムを検討されているから、彼らの結論は文を朗読する時の連続音声に対

して適応できるかどうかという疑問がある。本研究の目的は、人間が文を朗読するときの眼球運動と朗読音声を基づき、連続音声の発話計画メカニズムを推定することは目的にする。

## 1.2 本論文の構成

本論文は、6章で構成される。

### 第1章

この章では、音声生成過程の発話計画における研究の背景と問題点を述べながら、本研究の目的を明らかにする。

### 第2章

本研究は眼球運動と朗読音声を基づいた研究を行っている。そのため、この章は実験設計の全体像を紹介し、主に使った眼球運動指標と音声測定方法を説明する。そして、音声生成過程における発話計画の潜時時間と計画単位の分析方法を定義する。

### 第3章

この章では、前述な分析方法により、単純文、複雑文、複合文、無意味文を利用して行った三つの実験の結果から連続音声の発話計画潜時時間と計画単位の分析する。

### 第4章

この章では、第3章の結果に基づいた不足を述べ、脳電技術も加えて連続音声の発話を計画する時の脳活動を少し触って調査する。

### 第5章

この章では、本研究におけるテキストを朗読する実験で得られた発話計画の潜時時間と計画単位の考察を述べる。更に、その結果を従来研究と関連付けて全体考察を述べる。

### 第6章

本研究で得られた結果を結論として要約し、未解決な問題と今後の展望を述べる。

## 第2章 実験設計と分析方法

本研究では、従来の研究に使った孤立的な単語ではなく、一文ずつを被験者に朗読させる。実験の際に、被験者は文の中にある単語の情報を目で獲得しながら、朗読音声を放射する。そのため、文を見る時の眼球運動と音声を出す時の音声情報は必要である。この二つのデータを取得するために、以下のような実験を設計した。

### 2.1 実験の全体像

本実験は、主に眼球運動データと朗読音声データを利用されている。そのため、被験者が呈示した文を朗読する時の眼球運動は眼球運動装置（EyeLink 1000 Plus）で計測し、サンプリング周波数は1000Hzである。また、同時に計画した朗読音声はマイク（SONY ECM MS957）で記録し、サンプリング周波数は44100Hzである。実験の全体像は図 2.1 に示す。

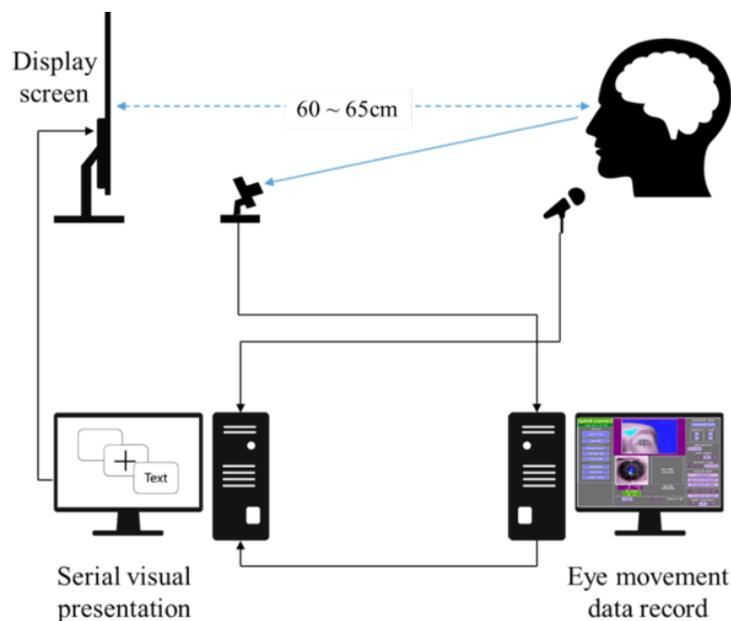


図 2.1: 実験の全体像

被験者側には、彼らの朗読音声を綺麗に取れるために、静かな部屋で行う。そして、被験者がディスプレイに向かって中央を見えるようにデスクの前に座り、目とスクリーンの距離は約 60cm (60cm から 65cm まで) に設け、眼球運動計測センサーは被験者の目線を邪魔しなくて、またデータを完全に取りれるようにディスプレイの間に設置する。また、参加者の頭の動きを防ぎ、額は固定する。実験を実施する前に、実験のプロセスと実験中に注意すべき点は被験者に説明する。

実施者側には、全ての実験は一台の host pc と眼球運動計測 pc でコントロールする。host pc において、MATLAB と PsychToolBox (version: 3.0.14) を利用してプログラミングした「眼球運動と音声測定システム」を運行し、実験を行う。刺激の呈示と朗読の音声データは host pc 側で集まる。また、このシステムは眼球運動計測 pc から探知した眼球運動データはリアルタイムに並行で獲得できる。だから、眼球運動と音声測定システムは実験を行いながら二つのデータを同時に収集する。

## 2.2 眼球運動と音声測定システム

眼球運動と音声測定システムのプロセスと実験の流れを次に紹介する。このシステムは二つの部分で構成している。一つは眼球運動校正であり、もう一つは実験のプロセスである。

### 2.2.1 眼球運動の校正

被験者の目線は文字の上に対応するため、本番の朗読実験を行う前に被験者の眼球運動の校正を行う必要がある。そのため、本実験は Eyelink にある 9 点校正機能を行い、注視精度の偏差が 0.50 以下になると合格である。図 2.2 と図 2.3 は合格な例を示す。



図 2.2: 眼球運動の校正

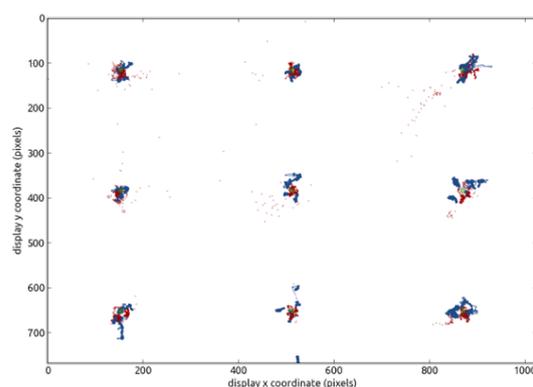


図 2.3: 眼球運動校正の結果

## 2.2.2 刺激呈示の流れ

目を校正したうえで，刺激提示の段階を行う．まず，センテンスを呈示する前に，”+”マックを用いて被験者の注意力を向上させる．持続時間は1000msである．そして，一つのセンテンスを被験者に読みさせ，眼球運動と朗読音声を計測し，情報をリアルタイムで保存する．この様な流れで，実験用コーパスのセンテンスをランダムで実験が行う．図2.4にセンテンスの呈示ごとに実験の流れを示す．

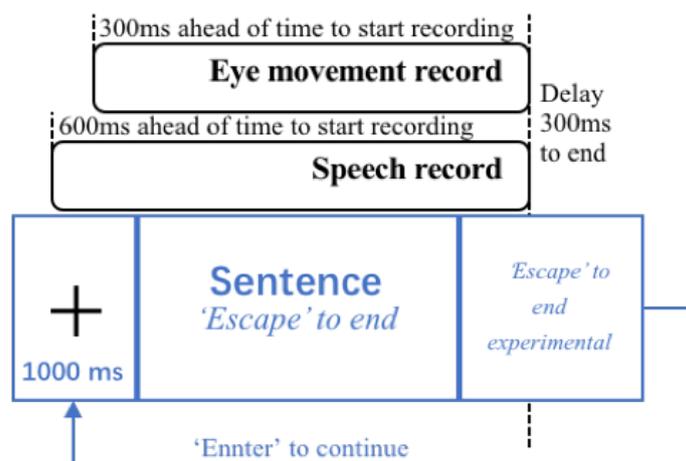


図 2.4: 実験の流れ

録音マイクと眼球運動記録機能の起動は550msと300msの遅延がある．その装置の開始遅延をなくなるため，「+」マークの呈示を開始後400msと700msの際に，記録装置がセンテンスを表す前，早めに起動する．次に，システムは文章を呈示し，朗読開始とデータ記録が始まる．最後に被験者の総朗読時間を把握するために，文章を読み終えた時，自らで「Esc」ボタンを押して記録を終了する必要がある．ボタンを押してから300ms後にマイクと眼球運動記録機能を停止し，データがデータベースに保存する．

## 2.3 潜時時間の測定方法

テキストを朗読する時の発話計画潜時時間とは，目線が単語を見る時刻から単語に離れた時刻までの時間帯である．その間には，目で獲得した単語の情報から意味を理解し，音素が重ね合わせて発話を計画することが発生している．そのため，この間の時間差は潜時時間に定義できる．

センテンスの中には多数な単語があり，各単語の潜時時間を切り出すために，眼球運動と音声の情報を同時に利用する．次には，今回の研究を果たすために利用された眼球運動

指標と利用方法について紹介する。また、連続音声データの各単語の発話音声を切り出されるツール及び方法を説明し、潜時時間がどのように計算するかという方法を述べる。

### 2.3.1 利用された眼球運動指標

従来の研究には、目線が目標単語に離れた時刻から発話開始時刻までの時間が発話計画の時間に計算された。しかし、連続音声の場合に、彼らの方法で計算した潜時時間は文末に近づくと、マイナスにする状況が出て来る。潜時時間はマイナスにすると正確に思わない。だから、発話計画は目標単語に注視する途中で開始すると考えられる。そのため、我々の潜時時間の測定については、三つの眼球運動指標が利用されている。以下にその三つの眼球運動指標を図 2.5 に示して各定義を説明する。



図 2.5: 眼球運動指標の図示

- ファスト注視点

目線が目標単語に入るはじめての注視点である。これは発話計画が始まる時、目標単語に注視が始まる位置を現われることである。図 2.5 の (1) (2) (3) (5) (6) (7) (8) (9) (10) (11) (12) 点はファスト凝視点の例で示している。

- ファスト注視時点

ファスト注視点の開始時刻である。これは潜時時間の開始時刻に定義する。

- 注視持続時間

目標単語を見る時の持続時間である。図 2.5 に示している注視点の大きさからなる。

眼球運動と音声の分析から測定する。まず、計測した眼球運動データは離線で被験者の目線移動がMATLABソフトウェアで再現する。再現した結果は図2.6に表している。赤い点は注視点であり、その位置はセンテンスを見る時の位置を描画している。注視点の大きさは注視持続時間を表す、時間が長いほどサイズが大きくなる。そのように被験者の朗読過程を再現するによって、注視開始時間から音声読み上げのオンセットまでの時間は潜時時間の計算に使用する。図2.7は再現した朗読の眼球運動軌跡を示す。縦軸は画面の左から右までの位置を示し、横軸は時間軸である。



図 2.6: MATLAB で眼球運動再現の結果 (例)

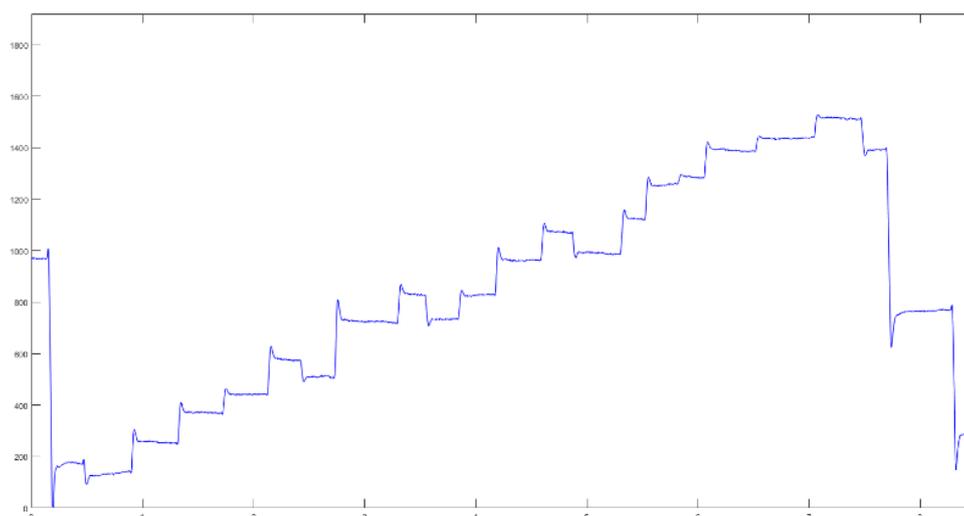


図 2.7: MATLAB で眼球運動軌跡の描画 (例)

### 2.3.2 各単語の音声切分

朗読音声は見たテキストを計画して生成した音声である。図2.8は一つの計画音声の結果を示している。時間軸に沿って各単語の音声は発話器官の運動によって生成する。をグ

ラムの軌跡をプロットしたものである。横軸は、そのスペクトログラムを背景とする連続音声の時間インデックスを表し、縦軸は、眼の動きを参照して単語境界を示す。

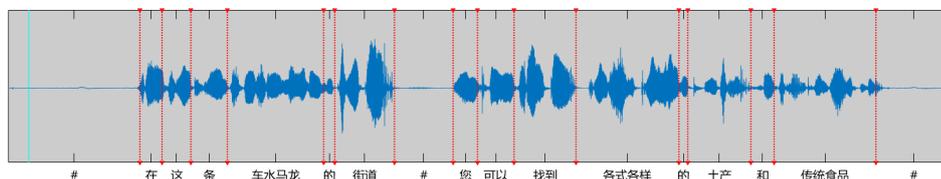


図 2.8: 朗読音声切分の結果 (例)

図 2.9 は、与えられたセンテンスに対する単語境界を有する音声の視線移動およびスペクトログラムの軌跡をプロットしたものである。横軸は、そのスペクトログラムを背景とする連続音声の時間インデックスを表し、縦軸は、眼の動きを参照して単語境界を示す。眼球運動は、2つの時間軸に沿ってプロットされる。各単語にどのくらい多くの注視点が分布しているか、どのくらいの長さの視線ポイントを直観的に観察して検証し、各単語の固定の開始点およびオフセットを見つけることができます。発症は、この単語の発話計画の開始タイミングとして定義される。読み上げ音については、連続音声の中の単語境界を得るために力アライメント法「」を用いた。図 4 に示すように、被験者が最初の単語「今年」(今年)を発音したとき、注視位置はすでに 2 番目の単語「除夕」(大晦日)に達し、後半の 3 番目の単語に移動しました最初の単語の発音の。その発話、音声の計画の潜在時間に応答する目の動きと比較して、大きな遅延を有することがわかる。

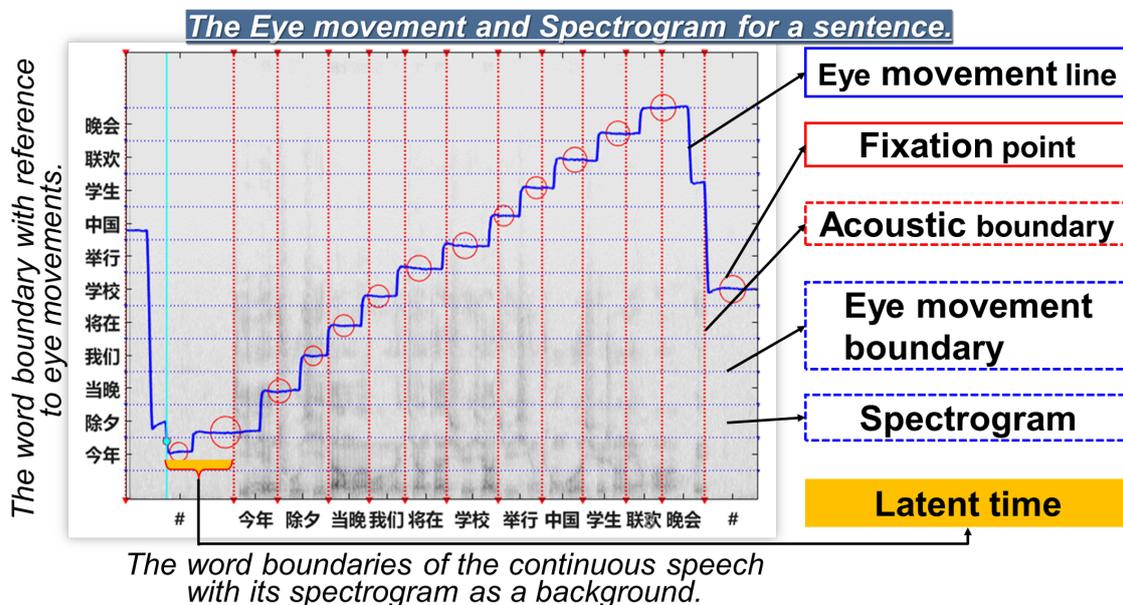


図 2.9: 眼球運動と音声測定の結果

## 2.4 計画単位の統計方法

人間がテキストを朗読する時に一つの注視点はどのぐらいの文字範囲で注視し、潜時時間を発生して発話しているということは計画単位とする。従来の眼球運動に着目していた閲読の研究には、その注視点の注視範囲を統計したことがある。中国語の文を閲読する時に、人間の目線が一文字から四文字までの範囲で分布し、多数な場合に二文字であることがわかった。眼動研究のような統計方法を利用して、読みながら発話する計画単位を分析する。

- 平均注視持続時間

一文を朗読する時に、全部の注視点の持続時間の平均値である。この眼動指標を利用して、文の中に有効的な注視点を選び出す。

- 総注視点数

初めての単語を注視開始の注視点から最後の単語を注視終わりまでの間にある注視点数である。

- 有効注視点

眼球運動パラメータに基づいて連続音声の場合の計画単位を分析する。Raynerらは、多数の言語を利用して黙読と音読の眼球運動を測定し、注視点の持続時間が200ms（黙読の平均注視持続時間＝225ms，音読の平均注視持続時間＝275ms）以上であることを見出した [6]。そのため、200ms以上の注視点は有効注視点と定義する。

## 第3章 データ分析と結果

本研究では、作成したセンテンスを朗読させ、眼球運動と音声測定を結合し、人間がテキストを朗読する時の発話計画メカニズムを解明することを目指す。そのため、三つの状況を考えられ、三回実験を実施された。まず、以前に証明された目の知覚広さの結果に従い、二文字単語だけを利用して単純文を作成する。それぞれの単語長は同じだから、連続音声の場合に、各単語の発話計画潜時時間はどのような風に変化があるかを探究する。

次に、長さが異なる単語を用いて二種類の複雑文を作成する。一つは単語の長さは、一文字、二文字、三文字または四文字からなる。もう一つは、異なる単語がのみならず、二つの短文でなっている長文である。これは単純文の結果を比較し、複雑文また日常的な文の場合になると、発話計画の潜時時間及び計画単位はどの様な影響があるかと調査する。

最後に、意味ある単純文と複雑文に異なり、単語の前後順位をランダムにして無意味の文を構築する。文の語彙またコンテキストは潜時時間に影響されるかどうかを検証する。

### 3.1 実験一：単純文の分析

過去の研究では、孤立的な単語を使用して人間の発話計画メカニズムを推測していた。しかし、単語ごとの発話計画メカニズムは本当の人間発話計画メカニズムに言えないと思われる。なぜなら、人間は多数の場合において連続音声で発話しているはずである。連続発話の発話計画メカニズムの解明を必要とする。そこで、連続音声における発話計画の潜時時間の変化を調べるために、単純的な中国語単語を用いてセンテンスを作成し、実験を設計する。

実験の参加者について、JAISTの大学院生10名（男6名、女4名）を募集して実験が実行した。すべての参加者の母語は中国普通語である。また、年齢は23歳から28歳の範囲になっている。眼球運動の情報をうまく取れるため、参加者は必ずコンタクトレンズやメガネを使用して、正常または矯正された視力を持っている。

#### 3.1.1 単純文の構築

単純文の構築については、眼球運動の視野範囲（知覚スパン）に基づいている。眼球運動に関する研究において、人間が朗読する時に、目の知覚スパンは1回の固視でほぼ二文字程度の情報を得られると報告されている。また、知覚スパンは読解材料の難易度に相関

するとわかれている。この二つの要件を満たすために、人間の目の知覚スパンに超えず、実験用コーパスはすべて二文字単語を用い、10個の中国語センテンスを作成する。センテンスごとに11個の二文字単語で構成される。そして、朗読する時に、内容の難易度に影響させないため、すべての単語は高頻度の単語が採用されている。コーパスは以下の図3.1に示す。テキストは被験者に提示する時に、すべての文字（大きさは35pt、形はMS明朝）はディスプレイの中央に1行で呈示されている。

1      2      3      4      5      6      7      8      9      10      11

|今年|除夕|当晚|我们|将在|学校|举行|中国|学生|联欢|晚会|

図 3.1: 単純文の構成 (例)

### 3.1.2 潜時時間の変化傾向

前述な潜時時間の測定方法に基づいて単純文の中に各二文字単語の潜時時間を計測した。結果は表3.1に現れているように、LT (latent time) は潜時時間であり、STD (standard deviation) は標準偏差である。

表 3.1: 単純文の平均潜時時間 (ms) と標準偏差 (ms)

単語位置	1	2	3	4	5	6	7	8	9	10	11
平均潜時時間	558	507	425	474	447	434	410	362	335	285	195
標準偏差	240	176	124	122	138	156	240	127	141	129	116

この結果から見ると、先頭単語の潜時時間は550ms (Avg = 558ms Std = 240ms) ほどなり、以前に明らかにした二音節単語の結果と一致することがわかる。しかし、二番目の単語から、対応の潜時時間は先頭単語に比べると、50msの差が出で来る。つまり、二番目の単語は発話を計画する時の時間が50ms (Avg = 51ms) に減少し、発話が早めている。また、次々の単語の潜時時間は先頭単語に比べると、多少の現象量があり、文末になる場合に、潜時時間は先頭単語の半分以下 (Avg = 195ms Std = 116ms) になっている。連続音声の場合に、先頭単語以後の単語は同じ二文字の単語にしても、計画時間が同じ程度ではない。

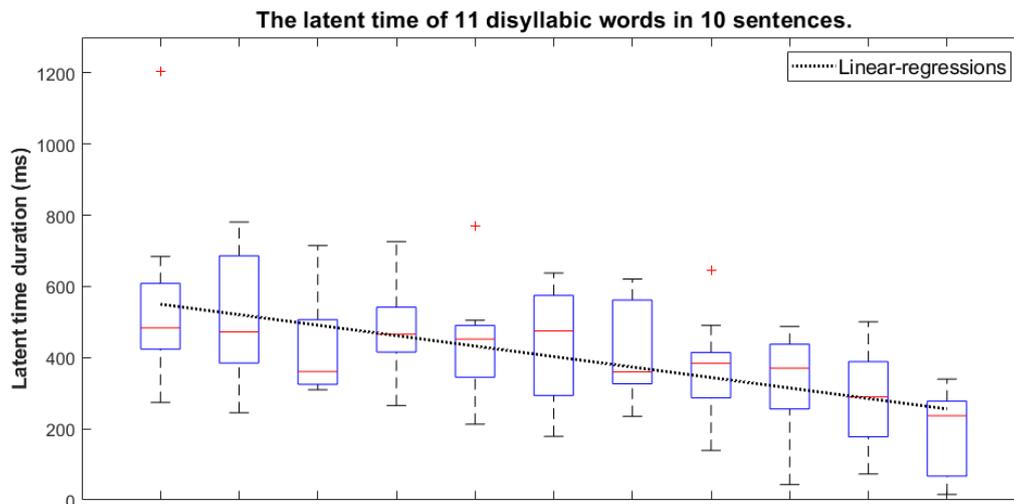


図 3.2: 単純文において各単語の潜時時間変化図

潜時時間の変化傾向は図 3.2 で表す。縦軸は潜時時間の持続時間であり，横軸は各単語の位置番号を表している。また，黒い線は連続音声の場合に各単語潜時時間の線形回帰した傾向を示している。傾向から見ると，二番目の単語から，その潜時時間は途中に登ったり下がったりしているが，文末まで 200ms (Avg = 195ms Std = 116ms) のところになっていた。先頭と文末には，300ms ぐらいの時間差があり，単純文が朗読する時に，発話計画の潜時時間は右下がりの傾向がある。

### 3.1.3 計画単位の統計

眼球運動パラメータに基づいて連続音声の場合の計画単位を分析する。Rayner ら (1998) は，多数の言語を利用して黙読と音読の眼球運動を測定し，注視点の持続時間が 200ms (黙読の平均注視持続時間 = 225ms, 音読の平均注視持続時間 = 275ms) 以上であることを見出した。我々の単純文朗読実験の眼動を再現して統計すると，平均注視持続時間は 256ms であり，標準偏差は 134ms である。これは Rayner ら (1998) の結論と一致している。そして，200ms 以上の注視点は有効注視点に採用することが合理的な判断であるとサポートできる。したがって，有効注視点は 200ms 以上の注視持続時間であると仮定して，発話計画の計画単位を明確にするために 200ms 未満の注視点を削除して計画単位を統計する。

先頭単語のファスト凝視点から文末単語のラスト凝視点までの間にある総注視点数を統計して分析した。その結果，単純文の場合に，有効注視点の数は 11 個 (Avg=11.57, Std=0.52) がある。それは，各単純文にある 11 個二文字単語の数とほぼ一致することに

なっている。つまり、一個の二文字単語の上に約一回の注視があり (Avg=1.05, Std=0.05), 単純文の朗読は二文字単語ごとで計画していると考えられる。

## 3.2 実験二：複雑文の分析

実験一に単純文の結果によれば、単語の位置が文末に移動するに従い、潜在時間が徐々に減少する。また、発話計画の計画単位は単純文の文法的単語と一致し、すべて二文字の単語からなることがわかった。しかしながら、二文字単語だけで構成された単純文は単語間の境界が多少明確し、人間目線が二文字で移動させることになる可能性がある。実環境にすると、文の朗読は二文字単語にのみならず、単語長が異なる場合は一般的である。もし文の単語長さが異なるにすると、潜在時間および計画単位両方が変わる可能性がある。もう一つの興味深い問題は、一つの文だけではなく、二つの短い文を使って一つの長い文にすると、潜時時間はどのような傾向がある。これらの理由、単語長が異なる多様な単語を用いて複雑な文章を作成し、発話計画の計画機構を調べる。

実験の参加者について、JAISTの大学院生20名(男性15名, 女性5名)を募集して実験が行った。すべての参加者の母語は中国普通語であり、年齢は23歳から28歳の範囲 (Avg = 24.3歳, Std = ...) になっている。すべての参加者は必ずコンタクトレンズやメガネを使用して正常または矯正された視力を持つ、眼球運動と発音テストを行って通過した。複雑文だから、事前に参加者の背景知識を質問して同じレベルになる。

### 3.2.1 複雑文の構築

今回の実験は、単語の長さが同じの制限をなくして、潜時時間と計画単位はどのような変化があるということを探る。そのため、実験一の単純文と違い、実験二には、二種類の複雑文を作成した。

#### 1. 複雑文

複雑文は、単語の長さが異なる高頻度単語でなっている。それぞれの文の長さが同じではないですけれども、単語の数が同じで、8個の単語がある。単語は、1音節から4音節までの単語である。全部7つの文が作成される。

1            2            3            4            5            6            7            8  
|五光十色|的|霓虹灯|把|城市|装扮得|更加|美丽|

図 3.3: 複雑文の構成 (例)

## 2. 複合文

複雑文は、単語の長さだけではなく、二つの短い文がある複合文である。それぞれの短文の長さが同じではないですけれども、単語の数が同じで、6個の単語がある。だから、一つの複合文は12個の単語がある。もちろん、短い文でも完全的な意味を持っている。全部15個の文が作成される。

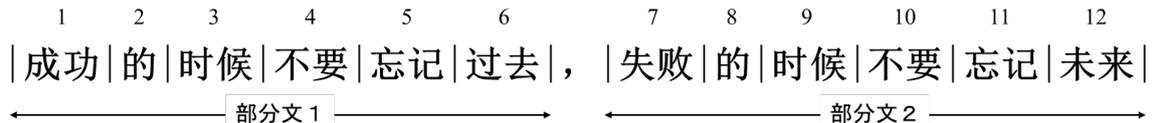


図 3.4: 複合文の構成 (例)

### 3.2.2 潜時時間の変化傾向

実験二では、実験一と同じ分析方法を使用し、潜在時間と計画単位を計測された。図 3.5 は、20 人の参加者に対する 7 つの複雑文 (類一) における各単語の潜在時間分布を示しており、黒い点線は潜在時間の線形回帰である。結果から見ると、文の中の単語が全く異なる音節数を有していても、実験一と同じ傾向が持っている。連続音声の潜在時間は右下がりの傾向がある。

表 3.2: 複雑文の平均潜時時間 (ms) と標準偏差 (ms)

単語位置	1	2	3	4	5	6	7	8
平均潜時時間	741	670	522	558	537	495	459	459
標準偏差	175	168	147	143	155	134	136	130

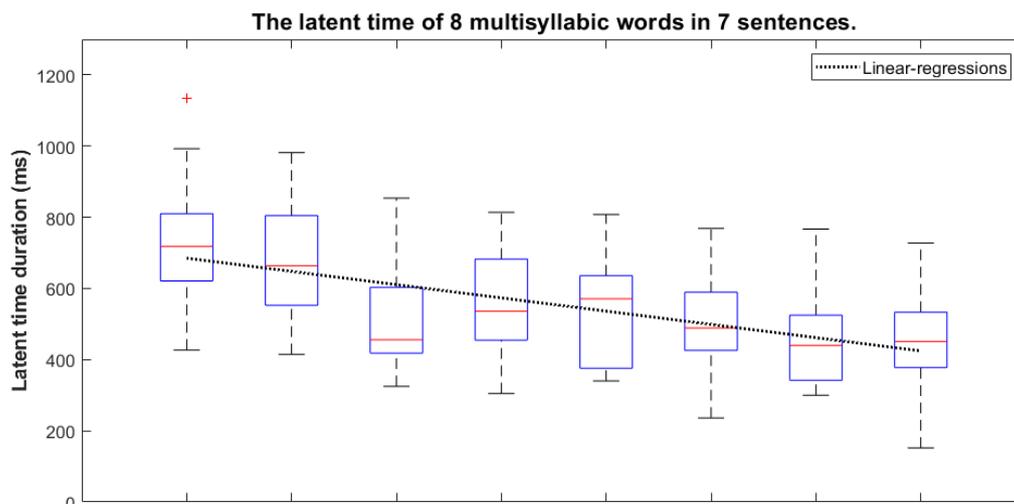


図 3.5: 複雑文において各単語の潜時時間変化図

図 3.6 は、2つの部分文から構成されていた複合文から測定した単語の潜時時間の分布を示している。縦の黒線は二つの部分文の間にあるコンマを示しているだけでなく、部分文間の境界でもある。結果から見ると、潜時時間は二番目の部分文に入るところにリセットされている。各部分文のうちにしても、右下がりの傾向は実験一の結果と一致している。

表 3.3: 複合文の平均潜時時間

単語位置	1	2	3	4	5	6	7	8	9	10	11	12
平均潜時時間	750	716	577	575	473	472	732	678	650	570	512	420
標準偏差	161	186	139	135	113	87	188	148	124	131	157	121

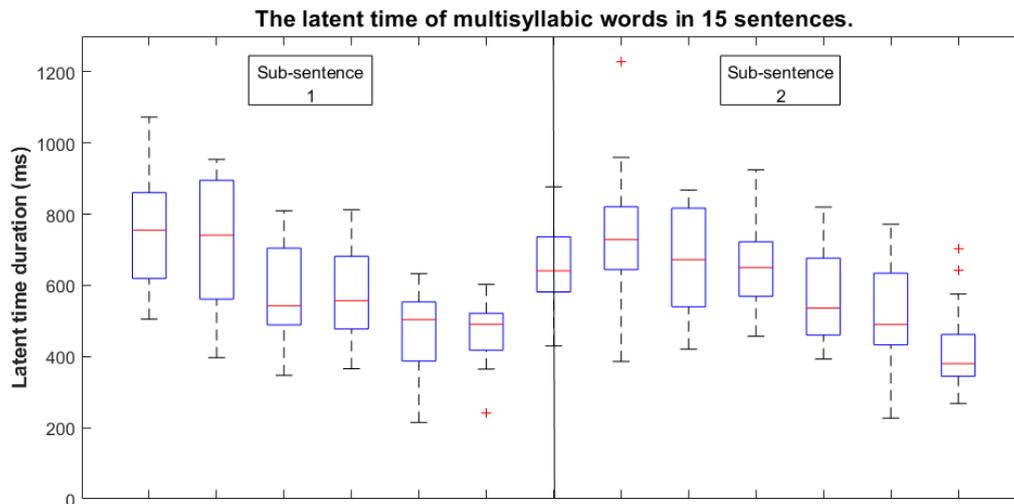


図 3.6: 複合文において各単語の潜時間変化図

連続音声の場合には、複雑文や複合文でも、単語の発話計画の潜時間は文の中の位置に強い関係があり、単語の長さや頻度の影響が見つからなくなった。つまり、前の単語を理解した上で、後ろの単語の理解に対して易くなった。前後単語の語彙が繋がっている。そのため、発話計画の潜時間は二つの部分になると考えられる。一つは、語彙理解であり、もう一つは発話運動指令生成である。

### 3.2.3 計画単位の統計

単純文ではなく、複雑文と複合文の発話計画単位を分析する。まず、実験コーパスの中にある文字長が一から四までの単語の数を統計した。それぞれの単語の分布は、図 3.7(a) に示されており、横軸は文法単語の音節数を表し、縦軸は文法単語の頻度分布を示している。この分布は、コーパス内の異なる音節を持つ単語の比率を反映している。統計した結果、一文字の単語は 23.8% で、二文字の単語は 50% で、三文字の単語は 13.6% で、四文字の単語は 12.6% になっている。

図 3.7(b) は計算した計画単位を示されており、横軸は計画単位の音節数を表し、縦軸は計画単位の頻度分布を示している。計算した結果、一つの注視点は一つの音節に注視している頻度分布は 33% で、一つの注視点は二つの音節に注視している頻度分布は 57.4% で、一つの注視点は三つの音節に注視している頻度分布は 7.4% で、一つの注視点は四つの音節に注視している頻度分布は 2.2% になっている。

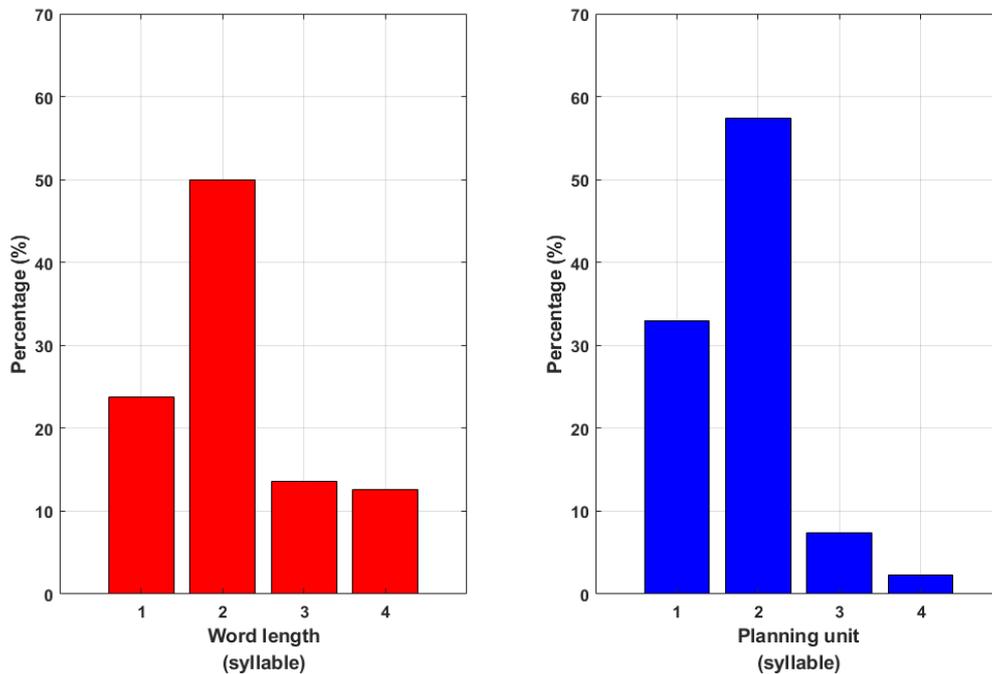


図 3.7: 複雑文と複合文の計画単位の統計

図 3.7(a) と (b) は対比してみると、単音節と二音節の単語は一つの単位として計画でき、三音節の単語は半分ぐらいになっている。四音節の単語はほとんど一つの計画単位に成れない。つまり、1 音節の単語は通常、1 つの計画単位として扱われる。二音節はコーパスよりもパーセンテージが高く、四音節の単語は左パネルよりもパーセンテージが低い。この現象は、計画段階で 4 音節の単語が 2 音節に分かれている可能性を示唆している。詳細を明らかにするために、我々は 4 音節の単語を分析し、それらの約 2/3 に 2 つ以上の固定点があることを見出した。この結果は、中国語における音声単位計画は基本的に単語に 3 つ以下の音節があるときは文法的な単語であることを示している。しかし、4 音節の文法的単語の大部分は一つの計画単位に成れない。また、半分ぐらいの 3 音節単語は一つの注視点で注視でき、計画単位として確定することはまたできない。

### 3.3 実験三：無意味文の考察

実験一および実験二の結果により、単語が文の後半部分に位置する場合、発話計画の潜時時間が短くなるという一般的傾向を示した。複雑な複合文では、その潜時時間は 2 つの部分文の境界で再設定されたが、各部分文の傾向は単純文の傾向と同じであった。連続音声では、単語の長さや頻度が潜在時間に与える影響は単語の位置が変動から影響と比べて、文の中の位置に強い関係があるとわかれていた。

連続音声における各単語の発話計画の潜時時間は、なぜ文の位置に強い関係があるか？我々は、潜時時間が語彙理解とモーター・コマンド設計という2つの部分で構成されていると推測する。もし、仮説が正しければ、次の単語の潜時時間が減少する原因は語彙理解の時間が短くなることに意味する可能性がある。この仮説を証明するために、我々は実験三を設計し、検証実験を行った。

### 3.3.1 無意味文の構築

我々は仮説を証明するために、文中の次の単語に対して前の単語の意味的支持を除外しなければならない。まず、実験一と二に使った単純文、複雑文と複合文のコーパスからそれぞれ三つを選択し、単語の順序を並べ替えて三種類の無意味文にする。

#### 1. 単純文 → 無意味単純文

単純文の場合には、各二文字単語だけの順位を並べ替えて、前後単語の意味が繋がらないで再編成する。全部12個の無意味単純文がある。

#### 2. 複雑文 → 無意味複雑文

複雑文の場合には、まず、四文字の単語を二つの二文字単語に分かれる。分かれた単語は、一、二、三文字の単語と同じ、順位を並べ替えて再配置する。各単語の意味は保持されて、全部12文の無意味複雑文を作成する。

#### 3. 複合文 → 無意味複合文

複合文の場合には、各部分文は複雑文と同じ方法で構築し、前後順位は変わらない。全部12個の無意味複合文を作成する。

総数36文の無意味文を作成し、実験が行った。

### 3.3.2 潜時時間傾向の変化

実験三では、実験一および実験二と同じ方法を使用して、無意味文を読む際に得られた眼球運動と朗読音声进行分析された。以下には、三つの無意味文から得られた結果を示す。

まず、無意味単純文における各単語の潜時時間の結果は表3.4に表している。結果から見ると、先頭単語の潜時時間は568ms (Std = 193ms) になり、意味ある単純文の結果と一致する。しかし、二番目単語から、各単語の発話計画潜時時間は上がったたり下がったり状況があるけれども、平均的に657ms (Std = 50ms) の持続時間になっていた。文末の単語も潜時時間は618ms (Std = 304ms) になり、先頭単語に比べると、少し増加になった。

表 3.4: 無意味単純文の平均潜時時間 (ms) と標準偏差 (ms)

単語位置	1	2	3	4	5	6	7	8	9	10	11	平均値
平均潜時時間	568	711	728	602	632	644	664	688	707	670	618	657
標準偏差	193	246	298	246	236	239	269	231	260	274	304	50

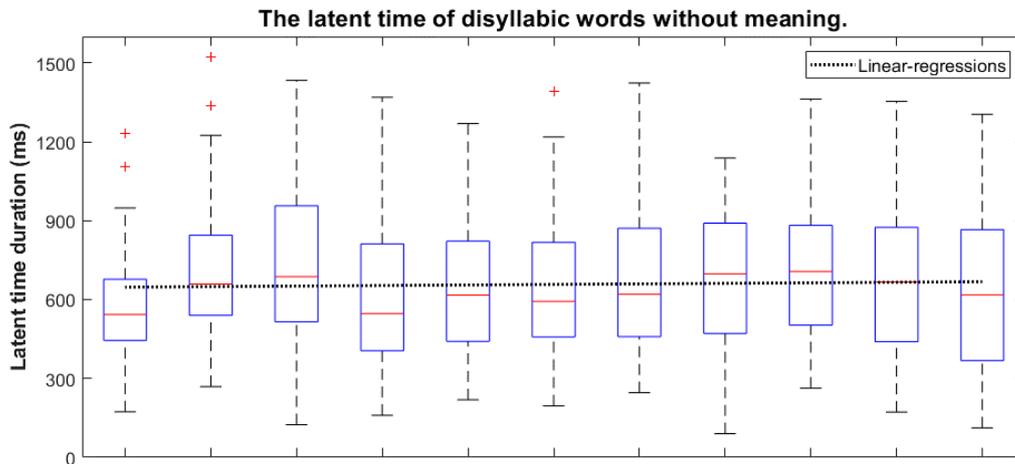


図 3.8: 無意味単純文の潜時時間変化図

そして、無意味複雑文における各単語の潜時時間の結果は表 3.5 に表している。無意味複雑文は単純文と結果大体同じく、平均的に 629ms (Std = 57ms) の持続時間になっていた。また、発話計画に対して単語長さや頻度の影響もなくなった。

表 3.5: 無意味複雑文の平均潜時時間 (ms) と標準偏差 (ms)

単語位置	1	2	3	4	5	6	7	8	9	平均値
平均潜時時間	570	744	693	618	578	632	601	640	585	629
標準偏差	248	275	322	272	245	254	232	238	245	57

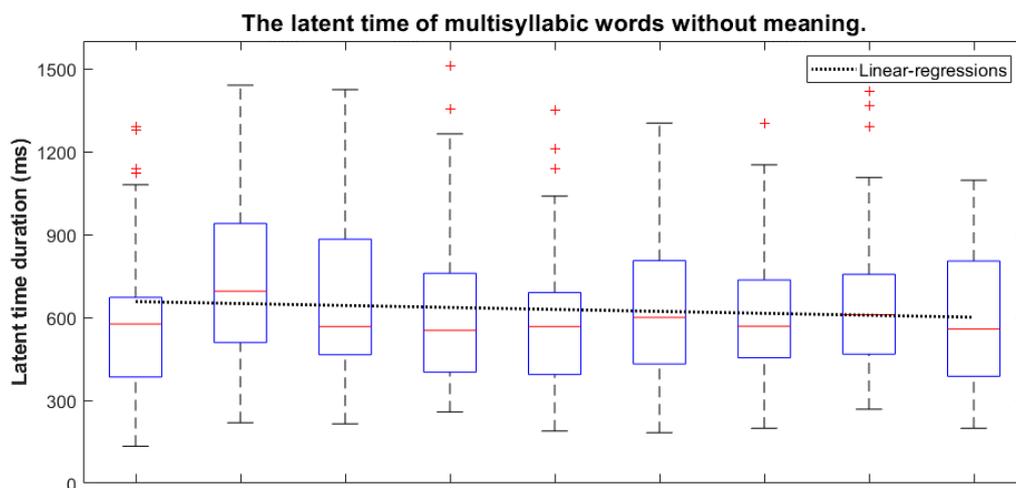


図 3.9: 無意味複雑文の潜時時間変化図

最後に、無意味複合文の場合にすると、二つの部分文の境界に潜時時間がりセットされなくて、潜時時間の変化傾向も消えた。つまり、それぞれの単語は、全部孤立的な単語として計画していると考えられる。

表 3.6: 無意味複合文の平均潜時時間 (ms) と標準偏差 (ms)

単語位置	1	2	3	4	5	6	7	8	9	10	11	12	13	平均値
平均潜時時間	603	698	709	642	623	563	717	704	748	705	648	622	598	660
標準偏差	215	250	204	209	233	203	243	224	230	249	227	240	177	57

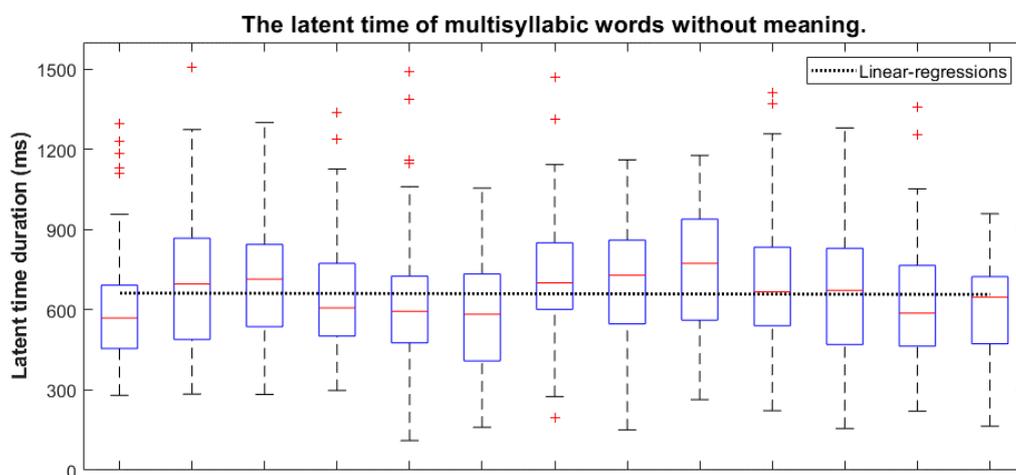


図 3.10: 無意味複合文の潜時時間変化図

以上に考察した無意味文の結果から見ると、一つのことを証明できた。それは、連続音声を計画する場合に、発話計画に対して、語彙理解の影響は単語の長さや頻度の影響より顕著することである。つまり、発話計画の潜時時間は二つの部分でなっている。語彙理解と発話運動計画である。

## 第4章 潜時時間における脳活動

テキストを朗読する際の音声生成過程は、おおざっぱに言うと、目で注視した単語を脳の視覚野を通し、角回で音とマッチングしてから、言語運動野で発話運動を計画して、発話・発声器官を制御しながら音声波を生成するというステップからなる [9]. これまで発話運動に伴う音声生成過程を脳電図による分析は困難であったため、テキスト音読の研究は、主に眼球運動指標に着目して、文字の注視から発話の開始までの反応時間（潜時時間）を測定しながら人間の発話計画メカニズムを推測してきた。眼球運動と音声測定を結合する先行研究では、文を用い、連続音声の場合の潜時時間を計測された。その結果、発話計画の潜時時間はその単語が文の中の位置に強い関係があり、前後単語の語彙理解の影響が顕著であることを明らかにした。しかし、潜時時間は発話計画する時の脳活動を側面から反映できると認められたが、脳における発話計画メカニズムの究明に対して、直接の証拠にならない。

眼球運動の時空情報を利用して

そのため、今回の実験は、眼球運動と音声測定だけではなく、脳電図も加えて連続音声の先頭単語の脳活動少し調査した。

### 4.1 実験の設計と流れ

#### 4.1.1 実験コーパスの構築

本実験では、音声コーパスは60個の中国語センテンスを作成し、四つのグループと分けられた。センテンスごとに8つの二文字単語で構成される。各グループには、一つの基本センテンスを設置し、学習させないため、単語が置き換えて類似なセンテンスを14個作成した。グループごとに15個のセンテンスがある。

#### 4.1.2 実験の設計と刺激呈示

本実験では、作成したセンテンスを朗読させ、眼球運動と脳電図（EEG）を結合し、音声測定も加え、人間がテキストを朗読する時の発話計画メカニズムを解明することを目指す。そのため、眼球運動は眼球運動装置（EyeLink 1000 Plus）で計測し、脳の活動は脳電計測装置（Neuroscan SynAmps RT 128-channel Amplifier）で収集する。眼球運動及

び EEG のサンプリング周波数が 1000Hz である．同時に計画した朗読音声も記録し，サンプリング周波数が 44100Hz である．図 4.1 に実験の全体像を示す．

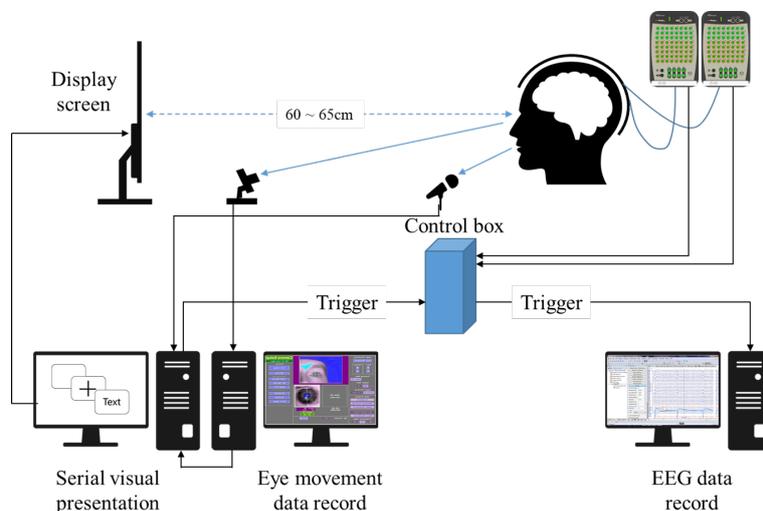


図 4.1: 眼球運動と脳電図に基づいた実験の全体像

刺激呈示には，まず眼球運動装置の校正を行う．被験者の脳活動をリセットするため，2000ms の空白画面を提示する．そして，センテンスを呈示する前に，“+” マックを用いて被験者の注意力を向上させる．持続時間は 1000ms である．この様な流れで，作成した 60 個のセンテンスをランダムで実験が行う．図 4.2 にセンテンスの呈示ごとに実験の流れを示す．

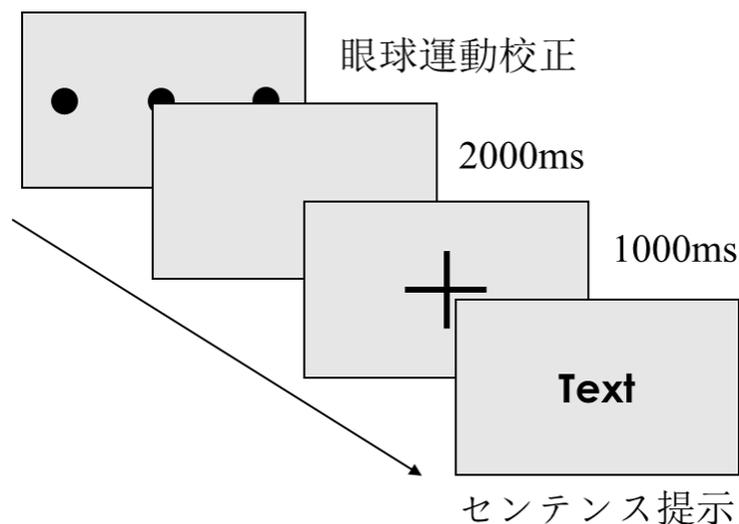


図 4.2: 刺激呈示の流れ

## 4.2 先頭単語の分析と結果

眼球運動のデータによって、先頭単語のファースト注視時刻（ORFt、目線が最初単語を見る時刻）とラスト注視時刻（OLFt、目線がはじめ単語に離れた時刻）は確定し、音声データによって、先頭単語の発話開始時刻（OUWt）は測定する。そして、三つのタイミングを利用し、先頭単語の注視持続時間（平均値：522ms）と脳活動の潜時時間（LT）を計算できる。潜時時間はORFtからOUWtまでの時間である。表4.1はコーパスの中に使用した四つの先頭単語の結果を示す。

表 4.1: 先頭単語の ORFt, OLFt, OUWt と LT の結果 (ms)

word	ORFt	OLFt	OUWt	LT
学校	357	1097	1171	814
天气	378	743	1081	703
中国	360	804	1087	727
境	332	872	1148	816
AVG	357	879	1122	765

そのタイミングによって、先頭単語だけの EEG 信号を切り出し、Eeglab (ver 14.1.1b) を用い分析する。主に使った方法は独立成分分析法 (ICA) である。ICA で得られた成分から眼動と雑音などの成分を取り除いて、有効な脳活動成分を求める。有効な脳成分から再構成した脳活動の領域を図4.3に示す。そこで、ORFt、OLFtとOUWtの区間において、脳活動の活発領域を赤のスポットで示す。スポットのサイズは活動の程度を比例している。

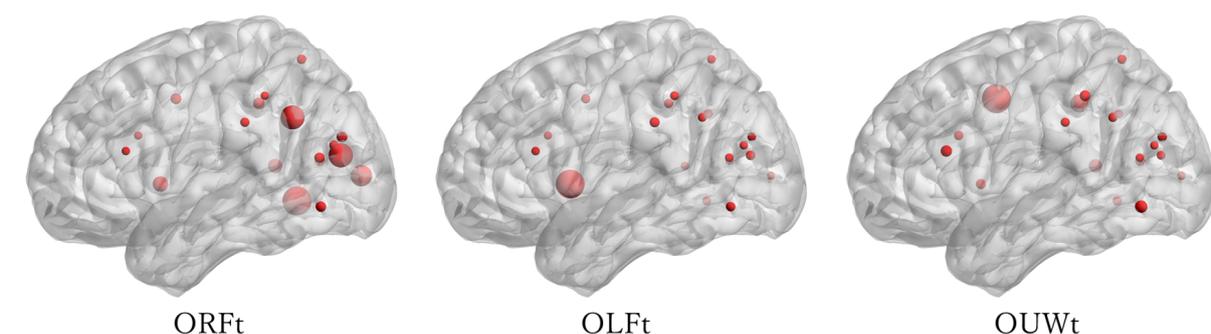


図 4.3: 先頭単語の脳機能活動図

脳の活動パターンから見ると、ORFtの区間では、最も活躍しているは視覚野（貢献率は28.4%）、角回（貢献率は27.8%）、ウェルニッケ野（貢献率は36.5%）である。これらの領域は、単語の認識、単語の形状と発音のマッチングおよび語意の理解を担当している。OLFtの区間では、ブローカ野（貢献率は21.6%）はほかの成分より活躍している。

ブローカ野は発話運動計画を担当しているため、発話計画をおもにこの期間で行われている。その後OUWtの時に運動野（貢献率は13.7%）に移行し、発話運動が始まる。その結果、眼動と発音の時期に概ねに合致している。それは、発話を計画する時の脳活動を反映することができると考えられる。また、Wernickeらが提唱された「言語の脳モデル」をサポートすることもできる。

図4.3に示したように、一つの機能を実現するため脳の複数の領域が一緒に活動する。異なる領域間のつながり・情報交換は、脳機能評価の常用指数である。そのため、上記の脳機能活動図に基づき、各期間における領域間の情報流れを加えて、テキスト朗読の脳ネットが構築された。その期間における脳のネットとその情報流れを図4.4に示す。

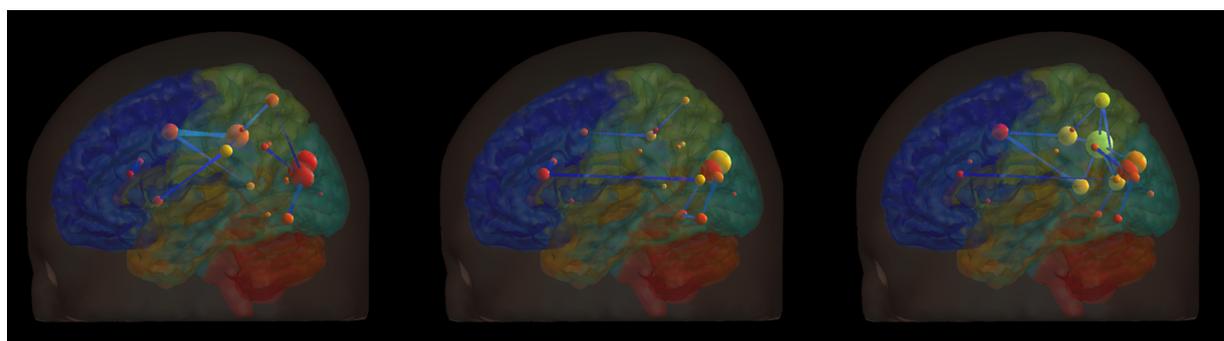


図 4.4: 脳ネットとその情報流れ

ノードは脳活動領域で、その大きさは活動の強さを表し、活動が強いほど色は赤くなる。エッジは領域間の情報流れを表している。情報はエッジを沿って、活動の強いノードに向かって流れる。この結果から見ると、潜時時間において、複数の脳領域が活発してお互いに情報を交換していることが分かる。

## 第5章 全体の考察

本研究では、眼球運動と朗読音声に基づいて人間が文を朗読する時の発話計画メカニズムを検討した。主なタスクは、連続音声における発話計画潜時時間の変化、および発話計画に使用される計画単位を分析することである。また、少し脳電図を加えて発話計画の脳機能と脳活動を調査した。

その結果、連続音声を潜時時間における単語の長さ起因する有意差を示さなかった。代わりに、潜時時間は文中の単語の位置に大きく依存していることがわかった。ここでは、文の後部に行くにつれて単語の潜時時間が直線的に減少する。文が複雑な文である場合にも、潜時時間の傾向は時間軸に沿って各部分文内で徐々に減少し、その減少傾向が2つの部分文の境界でリセットされた。

これらの結果に基づいて、潜時時間は語彙理解とモーター・コマンド設計という2つの部分で構成されていると推測できる。潜時時間が減少する原因は、先行する単語の語彙情報が次の単語の理解に助けがあるため、理解の時間を短縮することができる。つまり、前後単語の語彙が繋がっているため、後ろの単語の理解に費やされる時間が短縮される。このような仮説を証明するために、通常文の単語の順序を並べ替えて、次の単語の意味的支持を削除した。そのような無意味的な文を構築して検証したうえで、我々の仮説を証明した。

文を朗読するにおける発話計画の脳活動を解明するため、眼球運動と脳電図を併用して、潜時時間の脳活動を計測した。先頭単語の脳活動を着目して視線の移動と音声生成の脳活動に時間的に合致していることがわかった。また、文朗読時の脳ネットを構築し、脳の領域間の情報交換を明らかにした。

本研究では、理解課題を問わないが、発話課題において意味論的理解が自動的に伴っており、意味的理解が文章レベルまたは文章レベルで実行されることを意味する。文脈は意味論的理解のために重要です。複合文の結果は、意味論的不連続性が意味論的理解の活動をリセットすることを示している。無意味な文章を用いた実験では、単語の位置によって平均潜伏時間は変化しなかったが、その変化は通常文よりもはるかに大きかった。それは、意味のない文章の語彙情報を理解しようとする被験者もいるからである。

## 第6章 結論

### 6.1 本研究で明らかにされたことの要約

これまでの孤立単語の研究とは異なり、連続音声を用いて発話計画のメカニズムを調査した。先行する単語が増加するにつれて連続音声の発話計画の潜時時間は徐々に減少し、減少が過去の研究で主な要因として扱われた単語の長さによって引き起こされるよりも大きいことがわかった。潜時時間の短縮の起因を明らかにするため、それは語彙理解とモーターコマンド設計で二つの部分になることであった。無意味文の実験を行った結果、語彙理解には約 200ms で、全体的な発話計画の 3 分の 1 が必要であることがわかった。また、語彙理解は主に単語レベルではなく文脈レベルで行われることがわかった。そして、実験結果は、中国語では文法的な単語が基本的な発話計画単位であり、4 音節の単語の大部分は発話計画において一つの計画単位に成れない。

また、発話計画潜時時間の脳活動について、先頭単語の脳活動を着目して目線の移動と音声生成の脳活動に時間的に合致していることがわかった。また、テキスト朗読時の脳ネットを構築し、脳の領域間の情報交換を明らかにした。

### 6.2 今後の課題

これまで発話運動に伴う音声生成過程を脳電図による分析は困難であったため、テキスト音読の研究は、主に眼球運動指標に着目して、文字の注視から発話の開始までの反応時間（潜時時間）を測定しながら人間の発話計画メカニズムを推測してきた。本研究では、孤立的な単語ではなく、文を用い、連続音声の場合の潜時時間を計測された。その結果、発話計画の潜時時間はその単語が文の中の位置に強い関係があり、前後単語の語彙理解の影響が顕著であることを明らかにした。しかし、潜時時間は発話計画する時の脳活動を側面から反映できると認められたが、脳における発話計画メカニズムの究明に対して、直接の証拠にならない。そこで、朗読音声生成過程における発話計画に関する脳機能、脳活動また脳の情報流れのネットワークを解明する必要がある。

## 参考文献

- [1] Chiu, C. C., Phonological words in Mandarin speech production, In: Annual Meeting of the Berkeley Linguistics Society, Vol. 31, No. 1, pp. 61-72. pp. 61-72., 2005.
- [2] Meyer, A. S., Belke, E., Hcker, C., & Mortensen, L., Use of word length information in utterance planning, In Journal of Memory and Language, Vol. 57, No. 2, pp. 210-231., 2007.
- [3] Meyer, A. S., Roelofs, A., & Levelt, W. J., Word length effects in object naming: The role of a response criterion, Journal of Memory and Language, Vol. 48, No. 1, pp.131-147., 2003.
- [4] Kirov, C., & Wilson, C., Bayesian Speech Production: Evidence from Latency and Hyperarticulation., In CogSci.
- [5] Richardson, D. C., & Dale, R., Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension, Cognitive science, , Vol. 29, No. 6, pp.1045-1060., 2005.
- [6] Rayner, K., Eye movements in reading and information processing: 20 years of research., Psychological Bulletin, Vol 124, pp.372-422, 1998.
- [7] Yuan, J., & Liberman, M., Investigating consonant reduction in Mandarin Chinese with improved forced alignment, In: INTERSPEECH, pp. 2675-2678., 2015.
- [8] Levelt, W. J., Roelofs, A., & Meyer, A. S., A theory of lexical access in speech production., Behavioral and brain sciences, Vol. 22, No. 1, pp. 1-38., 1999.
- [9] 本多清志, 音声生成機構のモデル, 行動計量学, Vol. 22, No. 1, pp. 11-21., 1995.
- [10] Balota, D. A., & Chumbley, J. I., The locus of word-frequency effects in the pronunciation task: Lexical access and/or production?, Journal of Memory and Language, Vol. 24, No. 1, pp.89-106., 1985.