

話し手発話を基準とした発話開始位置の分析*

- 「はい」と「うん」を中心に -

○芦村 和幸 (JST/CREST) ニック・キャンベル (ATR, JST/CREST)

1 はじめに

前報 [1] で、話し手発話内に含まれる無音区間の有無により、聞き手発話開始位置が異なること、および発話 overlap の発生位置について報告したが、統計的分析としては必ずしも十分な量ではなかった。

本稿では、大量の対話音声データを利用し、話し手発話のモーラタイミングおよび発話内容にもとづいて、聞き手発話の開始位置を統計的に分析した結果について報告する。

2 音声試料

音声試料として、JST/CREST 発話様式プロジェクトで作成中の電話対話データベース [2] を用いた。

データベース中の音声に対しては、以下のような情報が付与されている。なお、各発話の開始・終了時間は、専門の作業員により、聴感上区切れが感じられる箇所においてマーキングされている。

- 発話内容の書き起こしテキスト (漢字かな混じり文)
- 発話開始・終了の時間情報
- 談話行為タグ (質問, 応答, あいづち, など)
- 感情・態度タグ

データベースのうち、日本語を母語とする話者 6 名による対話音声进行分析対象とした。表 1 に対話の組み合わせと収録件数を示す。

表 1 対話の組み合わせと収録件数

話し手	聞き手						合計
	JFA	JFB	JFC	JMA	JMB	JMC	
JFA	-	10	-	10	-	-	20
JFB	10	-	10	-	-	10	30
JFC	-	10	-	-	11	-	21
JMA	10	-	-	-	11	-	21
JMB	-	-	11	11	-	10	32
JMC	-	10	-	-	10	-	20
-	-	-	-	-	-	-	144

※注意: 1 対話の収録時間は 30 分である。

3 発話単位

朗読音声の無音区間長に対しては、平均モーラ長にもとづくモーラタイミングが影響することが報告されている [3]。自由対話の発話開始位置に対してもモーラタイミングが影響するかどうかを検証するために、「話し手発話開始時点から、次の話し手発話開始時点までの区間」を分析のための発話単位とし、話し手発話開始からのモーラ位置を基準として、聞き手発話の開始位置を分析する (図 1)。

なお、発話 overlap がない場合は、聞き手発話が無音区間で開始されるため、対応する話し手発話のモーラが存在しないが、話し手発話完了から聞き手発話開始までの無音継続時間 (以下、発話間ギャップ) は話し手発話のモーラタイミングにより影響を受けている

*Analysis of utterance beginning position based on addressee's utterance - using Japanese frequent expressions /hai/ and /uN/, by Kazuyuki ASHIMURA (JST/CREST) and Nick CAMPBELL (ATR, JST/CREST)

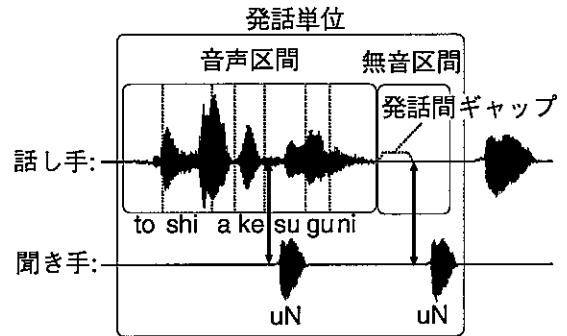


図 1 話し手発話を基準とする聞き手発話開始位置と考へ、話し手発話音声区間の平均モーラ長により発話間ギャップ長を正規化する。

4 モーラ情報の取得方法

各発話を構成するモーラごとの時間情報を得るために、まず、音声、書き起こしテキスト、および発話単位の時間情報にもとづいて全対話データから 115,154 発話を切出した上で、話し手発話単位内で聞き手発話開始があるもの 79,010 件を抽出した。次に、各発話に対して、日本語形態素解析システム茶筌 [6, 7] により読み付与を行なった上で、大語彙連続音声認識エンジン Julius [6, 8, 9] により音素セグメンテーションを行ない、その結果からモーラの時間情報を算出した。

以下、音素セグメンテーションが可能であった発話 68,753 件を対象とした聞き手発話開始位置の分析結果を示す。

5 分析結果と考察

68,753 件の話し手発話に対して 100,012 件の聞き手発話開始がみられたが、聞き手発話の書き起こしテキストには 58,295 種類ものバリエーションがあり、全バリエーションについて分析を行なうのは不可能なため、特に利用頻度の高い /hai/ および /uN/ について分析を行なった。表 2 に、聞き手発話における /hai/ と /uN/ の利用頻度を示す。

表 2 /hai/ と /uN/ の利用頻度

発話 overlap の有無	/hai/	/uN/
発話 overlap なし	1,700	1,295
発話 overlap あり	2,084	2,428
全体	3,784	3,723

以下、主に /hai/ の分析結果を示すが、/uN/ の分析結果も同様であった。

5.1 発話 overlap がない場合

朗読音声の無音区間長は、時間の対数軸上での正規確率密度関数で比較的良好に近似できることが報告されている [4, 5]。対話音声における発話間ギャップについても同様の傾向が見られるか確認するため、発話間ギャップ長の、対数軸上での確率密度関数を求めた (図 2)。その際、話し手のモーラタイミングによる影響

を考慮し、発話間ギャップ長を話し手発話の平均モーラ長で正規化した上で対数化した¹。

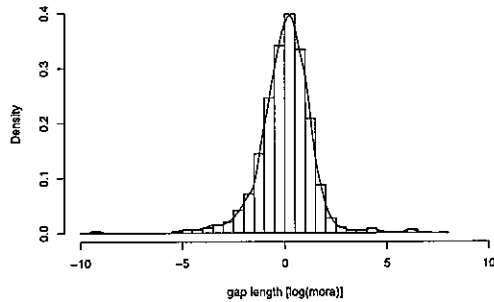


図2 発話間ギャップ長の分布と確率密度関数

発話間ギャップ長の確率密度関数は、正規確率密度関数とよく近似しており、表3に示す分布の統計量より、発話間ギャップは1モーラ程度の場合が最も多いといえる。

表3 発話間ギャップ長の分布に関する統計量

統計量	対数モーラ値	モーラ換算値
平均	0.05	1.05
標準偏差	1.33	-
レンジ	-9.21 ~ 6.54	0 ~ 690
95% 信頼区間	-1.23 ~ 1.33	0.3 ~ 3.8

5.2 発話 overlap がある場合

話し手発話先頭からのモーラ数により計測した発話 overlap 発生位置と、各位置での発話 overlap 発生頻度を調べた上で、モーラ数の対数軸上での確率密度関数を求めた² (図3)。

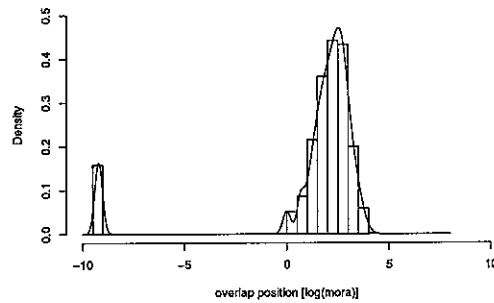


図3 overlap 発生位置の分布と確率密度関数

図3で、対数モーラ軸上の2近辺にピークを持つ分布の確率密度関数は、正規確率密度関数とよく近似しており、表4に示す分布の統計量から、発話 overlap 開始位置は9モーラ付近である場合が最も多いといえる。なお、-10近辺にピークを持つ分布は「話し手発話開始と同時の聞き手発話開始」であり、/hai/では64件、/uN/では212件みられたが、統計量には含まなかった。

表4 overlap 発生位置の分布に関する統計量

統計量	対数モーラ値	モーラ換算値
平均	2.19	9
標準偏差	0.81	-
レンジ	-0.0001 ~ 4.39	1 ~ 81
95% 信頼区間	1.41 ~ 2.97	4 ~ 19

前報 [1] で、話し手発話末尾における overlap が多いことを報告した。しかし、表4に示す通り overlap 発生位置分布は平均 2.19 (9モーラ)、標準偏差 0.81

¹ 発話間ギャップが0となる場合を考慮し、モーラ単位に正規化した発話間ギャップ長に0.0001モーラを加えた上で対数化した。
² モーラ位置が0となる場合を考慮し、モーラ位置に0.0001モーラを加えた上で対数化した。

である一方、overlap がある場合の話し手発話モーラ長分布は平均 2.64 (14モーラ)、標準偏差 0.72 であり、発話末尾における overlap 発生は確かに多そうではあるものの、統計的分析結果から明確に結論づけることはできない。そこで、聞き手発話 /hai/ および /uN/ による発話 overlap が発生する箇所全てについて、overlap 直前の話し手発話内容 4モーラ (以下、発話コンテキスト) を確認し、前報 [1] で挙げた発話 overlap 発生位置ごとに分類した (表5)。

表5 発話コンテキストの出現頻度

発話 overlap 発生位置	種類数	合計件数	割合
聞き手の /hai/ の直前			
発話末尾付近	234	1353	65%
母音の引き延ばし	624	1022	49%
発話先頭付近	262	458	22%
/su/ の後	65	119	6%
聞き手の /uN/ の直前			
発話末尾付近	229	1430	59%
母音の引き延ばし	693	1091	45%
発話先頭付近	328	587	24%
/su/ の後	123	294	12%

※注意: 種類数は、各発話 overlap 発生位置に分類された発話コンテキストのパリエーション数。合計件数は、各分類ごとの、発話コンテキスト出現頻度の合計。割合は、/hai/ による overlap 2084件、および/uN/ による overlap 2428件に対する割合。なお、複数の条件を同時に満たす場合があるため、割合の総計は100%より多い。

表5から、大量データによる分析においても、「発話末尾付近」、「母音の引き延ばし」および「発話先頭付近」で発話 overlap が発生することが多いことが確認された。

6 おわりに

大量の対話データを利用し、話し手発話のモーラタイミングおよび発話内容にもとづいて、聞き手発話の開始位置を統計的に分析した結果について報告した。今後は、音素表記やモーラ数のみならず、談話行為タグ、品詞、係受け関係および韻律情報などの利用について検討を進めるとともに、発話単位中に聞き手発話の開始が見られなかった場合についても分析を行ない、発話開始位置設定の規則化を目指す。

謝辞

いつも有益なアドバイスをくださる、ATR-SLTの柏岡秀紀主任研究員、山本博史主任研究員、および本プロジェクトのパーハム・モクタリ研究員、カルロス石井研究員に感謝いたします。

参考文献

- [1] 芦村, キャンベル: 発話速度を考慮した発話タイミング分析, 音講論 秋季 3-10-10, pp.347-348 (2002).
- [2] 芦村, キャンベル: JST/CREST 発話様式プロジェクトの電話対話データベース, 人工知能学会全国大会論文集 3C5-11 (2002).
- [3] 海木, 勾坂: 局所的な句構造によるポーズ挿入規則化の検討, 信学会論文誌 D-II, Vol. J79-D-II No.9, pp.1455-1463 (1996).
- [4] 藤崎, 大野, 山田: 朗読音声における句間休止の長さに対する各種要因の影響, 音講論 秋季 2-2-11, pp.245-246 (1997).
- [5] 藤崎, 大野, 山田: 文章朗読における発話と休止の時間的特徴, 音講論 春季 1-7-5, pp.209-210 (1997).
- [6] 鹿野, 伊藤, 河原, 武田, 山本: 音声認識システム, IT Text シリーズ, オーム社 (2001).
- [7] 松本, 北内, 山下, 平野, 松田, 高岡, 浅原: 形態素解析システム「茶釜」 version 2.3.0 使用説明書, <http://chasen.aist-nara.ac.jp/> (2003).
- [8] 米良, 李, 猿渡, 鹿野: Julius を用いた自由発話の自動ラベリングにおけるラベル尤度の統計的分析, 音講論 秋季 2-1-5, pp.57-58 (2001).
- [9] 李, 米良, 鹿野: Julius によるマルチパス音韻モデルを用いた対話音声認識, 第2回 話し言葉の科学と工学ワークショップ講演予稿集, pp.121-128 (2002).