

「見る」と「聞く」の言語理解の観察 -いい加減なコミュニケーションの仕組み構築に向けて-

谷田泰郎^{†1} 高椋琴美^{†1}

人の理解はいい加減なものである。本稿では、人の理解や記憶のいい加減さに着目し、小説の朗読音声「聞いた」場合とテキストを「見た」場合での記憶の違いを比較し、言語特徴や音響特徴から機械的に抽出したものと人間が記憶して書き起こしたものを要約として評価することで、いい加減なコミュニケーションの適用可能性について探る。

The observations on language understanding of the "read" and "hear" -Toward the system architecture on the sloppiness of human communication-

YASUO TANIDA^{†1} KOTOMI TAKAMUKU^{†1}

The human understanding is sloppy. In this paper, we focus on the sloppiness of human understanding and memory. And we compare the storage difference of "read" that "hear". By evaluating the place extracted from the language features and acoustic features and the place extracted from human, We explore the applicability of sloppy communication.

1. はじめに

人間は多くの時間を1対1のコミュニケーションに費やしており、1日のうち6~12時間を知り合いと1対1で対話し、そのうちの80-90%を世間話に費やしているという[1, 2]。つまり、コミュニケーションは社会的グルーミングであり、武器でもあり、社会を生き抜くための練習だともいえる。生物は生得的に生きているものを知っていて自分と同じ種だと言うことが分かるというが、私たちは人の話を聞いたり様子を見て心が傷んだり、心地よくなったりするだけではなく、非生物的なもの（例えばぬいぐるみや自分が映っている鏡）に対してさえ擬人化して対話をする。学べないものに対して教えるという行為の中で自分自身が学んでいることもある。コミュニケーションに欠かせない言語は他者への情報伝達のためだけにあるのではない。なぜなら、人間のコミュニケーションがいい加減だからである。筆者らは、コミュニケーションの中から「心のモデル」をデザインするためのエビデンスを収集したいと考えている。その中の取り組みの一つが「いい加減な」対話からの情報抽出である。

手始めに、言語というインターフェースを介したコミュニケーションの「いい加減さ」に着目し、小説のテキストを見るという行為と朗読を聴くという行為で得られる情報を人がどのように解釈し、それぞれの入手方法によってどのような違いがあるのかを知るための実験を行った[3, 4]。本報告では、その実験の内容を紹介するとともに、得られた知見をどのような技術に応用していくのかについて示唆する。

2. 過去に行った視聴覚実験

視覚と聴覚の文書呈示による、あるいはもっと一般的な記憶の違いや活動の違いを整理しようとしている研究[5, 6, 7]は数多くあるが、決定的な結論には至っておらず、我々視覚と聴覚の様々な違いに関する一般的な知識はあるが、その感覚が言語を中継した場合に、どのような差となって現れるのかについてはよく分かっていない。

また、人間は経験学習を通じて体得している自然言語フレームを活用して入力刺激を要約して解釈していると考えられる。その詳細を知るのは困難であるが、少なくとも入手経路（視覚と聴覚）に応じて計算コストのかからない方法で重要な部分を抽出しようとしているはずで、その様子を観察することはできる。そこで、それらを確認するために、被験者に小説のテキストを読んだり、朗読を聞いたりしてもらい、その記憶に残ったものを書き起こしてもらいという簡単な視聴覚実験を行い、テキストを入力にした場合と朗読を入力にした場合でどのような違いがあるのか、男女の違いはあるのかなどを定量的、定性的に観察した。

2.1 視聴覚実験の概要

具体的には、夏目漱石の小説「ころ」(「上三」の章の前半4段落及び「下四十五」の章の前半3段落)の朗読音声「聞いた」場合とテキストを「見た」場合での意味理解の過程を記録してもらった。なるべく単語でなく文章にして他人にあらすじを説明するようつもりで書くように指示した。

表1に示すように、全被験者21人(男性11人、女性10人、年齢は19歳から52歳までの平均29才、題材の小説を知らない、読んだことのない人を対象)を午前と午後の2

^{†1} シナジーマーケティング(株)
Synergy Marketing, Inc.

部に分け、午前の部は「上三」をテキスト、「下四十五」を朗読、午後の部は「下四十五」をテキスト、「上三」を朗読というようにテキストと音声の実験素材を午前と午後で逆にする事で、同じ素材でのテキストと朗読音声の実験結果の比較ができるようにした。午前、午後ともテキスト実験、朗読実験の順に行った。テキストの見取り、朗読の聞き取り実験をそれぞれ3回ずつ行い、1回ごとに記憶している内容を書き起こしてもらった。2回目以降はそれ以前に自分が書き起こしたテキストを見ずに作業してもらった。書き起こしテキストとともに、1回ごとに体感的な指標として、どれぐらい理解できたか、どれぐらいイメージを書き出せたか、その回の作業に対するコメント・感想を聴取した。書き起こし作業の時間は最大20分で設定し、全員の書き起こし作業が終わったところで切り上げた。表に示したように1回目が11分程度、2回目が15分程度、3回目が18分程度で終了している。また、朗読音声の長さが3~4分であったため、テキストを読む時間を3分とした。これらの書き起こし実験回数や時間設定は予備実験を3人に対して行った体感値で定性的に決定した。

表1 実験の条件

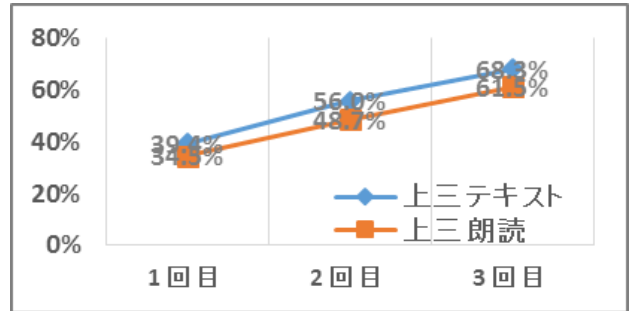
	被験者数	実験テキスト	記憶の書き起こしに要した時間	記憶の書き起こしに要した時間	実験朗読音声	記憶の書き起こしに要した時間
午前	11人	上三	1回目	11.5分	下四十五	11分
			2回目	15分		15分
			3回目	18分		17分
午後	10人	下四十五	1回目	11分	上三	11分
			2回目	15分		15分
			3回目	18.5分		18分

また、実験で使った素材に関しては、「上三」は、主人公である「私」と「先生」が海水浴場で話す機会を得るまでの話で風景描写などの叙事的な記述が多く、「下四十五」は若いころの「先生」が「K」という友人を出し抜いて「お嬢さん」と結婚するためにその母親である「奥さん」に結婚を願い出る場面で会話が多く感情的な記述が多い、というように内容としては対照的な素材を選択した。

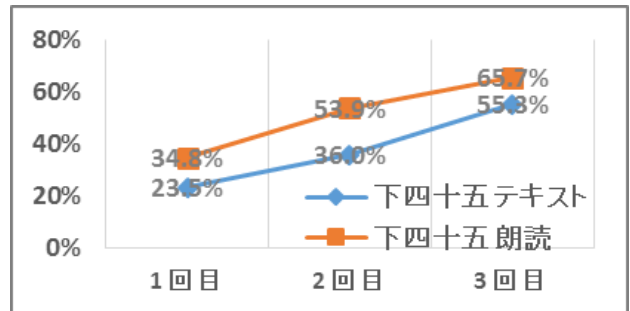
2.2 実験結果の定量的観察

テキストや朗読の実験対象が変わっても、難易度の個人差こそあれ何度か聞くうちに徐々に書き込み量は増えるはずである。然しながら、その増加内容には、実験対象による違いや男女による違いがあると考え、記述量の変化を調べた。グラフ1とグラフ2は、「上三」及び「下四十五」を実験対象とした時の、テキスト実験及び朗読実験における記述量の変化である。記述量を分かりやすく提示するために、元のテキストの形態素数（「上三」が569形態素、「下四十五」が598形態素であった）に対して被験者が書き起こした形態素数の割合（被験者平均）をグラフの数字とした。

段階的な記憶の再現割合は「上三」がほぼ線形的、「下四十五」は2回目から3回目の再現割合が朗読の場合は小さくテキストの場合は大きい。そして、「上三」がテキストの場合の再現割合が多いのに比べ、「下四十五」では朗読の場合の再現割合が多い（その差も大きい）。このように、定量的な再現割合では、対象の内容が叙事的であればテキスト優位、感情的であれば朗読優位という結果になった。



グラフ1 「上三」における記述量の変化



グラフ2 「下四十五」における記述量の変化

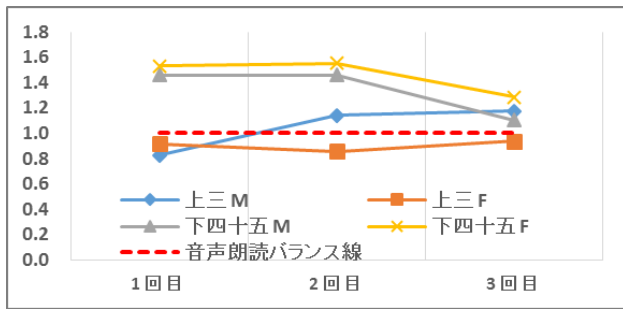
表2は各実験条件における男女別のテキスト及び朗読に対する記述量（再現割合）の変化である。男女別では、全体的に女性の方が男性より記述量が多い。テキストと朗読を比べるとテキストより朗読の方がその差が大きくなる（テキストでは男女差が殆どないにも拘らず、朗読では女性の記述量が増える）傾向がある。

表2 男女別の記述量

実験条件		sex	1回目	2回目	3回目
上三	テキスト	M	38.7%	53.2%	65.8%
		F	40.4%	59.5%	71.2%
	朗読	M	32.0%	46.5%	56.0%
		F	37.0%	51.0%	67.0%
下四十五	テキスト	M	21.2%	33.9%	56.6%
		F	25.8%	38.1%	54.1%
	朗読	M	31.0%	49.5%	62.3%
		F	39.5%	59.2%	69.7%

グラフ3に朗読実験の再現割合をテキスト実験の再現割合で割った倍率のグラフを示す。1.0倍で朗読・テキスト差がないということになる。回数を重ねるごとに朗読もテキストも1.0に向かって収束していくはずである。「上三」

では男女とも差が小さいが、「下四十五」では男女とも朗読優位の結果が出ており、特に女性ではその傾向が顕著なる。



グラフ 3 朗読再現割合/テキスト再現割合

2.3 実験結果の定性的観察

定性的な評価の指標として、テキスト及び朗読の各実験において被験者が小説の内容をどれぐらい理解して書き起こしているかを調べた。「上三」及び「下四十五」のテキストを意味のある単位（文章より短く、単文の単位に近い）に恣意的に分割し、全被験者のテキスト及び朗読の1回目の各実験（最初の1回目に「いい加減」に理解するという知見が多く現れると考えた）で書き起こされたテキストにその内容を表す重要なキーワードが含まれているかを検証した。キーワードが含まれていれば記憶が再現したとみなして再現割合を求め、定性的な評価指標とした。以下に、意味単位の区切り例と表3に再現率の高かった（多くの被験者が記憶していた）意味単位を示す。

【意味単位の区切り例】

私が「急に貰いたいのだ」とすぐ答えたら／(区切り)
 笑い出しました。／(区切り)

表 3 再現率の高かった意味単位

優位	分類	意味単位	段落
朗読優位	会話部分	「下さい、ぜひ下さい」といいました。	下四十五 結婚申込み
		私が「急に貰いたいのだ」とすぐ答えたら	
	音声で強調	「差し上げるなんて威張った口の利ける境遇ではありません。ご存じの通り父親のない憐れな子です」と	下四十五 結婚承諾
テキスト優位	叙述的な記述	男のように判然したところのある奥さんは、普通の女と違って最初からしまいまでにおそらく十五分とは掛らなかったでしょう	下四十五 結婚承諾
		その上先生の態度はむしろ非社会的であった。	上三 海での先生の様子
		すると着物の下に置いてあった眼鏡が	上三 眼鏡の紛失
女性のみ記憶	比較強い体質をもった私は、	Kから聞かされた打ち明け話を、	下四十五 結婚申込み
		奥さんに伝える気のなかった私は、	
男性のみ記憶	板の隙間から	「愉快ですね」と私は大きな声を出した。	上三 先生と海へ
		「帰りましょう」の段落	上三 帰りましょう

表3に示すように、意味のある単位に区切った評価を一つずつ調べたところ、朗読優位で記憶に残っていたものは、「差し上げるなんて威張った口の利ける境遇ではありません」「ご存じの通り父親のない憐れな子です」と（「下四十五」「結婚承諾」の段落）「下さい、ぜひ下さい」といいました。「私が「急に貰いたいのだ」とすぐ答えたら」（「下四十五」「結婚の申し込み」の段落）のような会話の部分が多かった。また、「男のように判然したところのある奥さんは、普通の女と違って」（「下四十五」「結婚承諾」の段落）

“最初からしまいまでにおそらく十五分とは掛らなかったでしょう。”（「下四十五」「周囲・本人への承諾」の段落）のように朗読音声で強調していたと思われる箇所もあった。テキスト優位であったものは、「上三」の「海での先生の様子」の段落や同じく「上三」の「眼鏡の紛失」の段落のように叙述的な記述が続く部分であった。また、「下四十五」の「結婚申し込み」の段落の「Kから聞かされた打ち明け話を、」「奥さんに伝える気のなかった私は、」については朗読では印象に残らなかったのかテキスト優位であった。被験者の書き起こしを見た感想としては、朗読実験の方が自分の言葉で書き起こしていた。テキストより朗読の方が言葉の置き換えが起こりやすいものと考えられる。

その他、「愉快ですね」と私は大きな声を出した。（「上三」「先生と海へ」の段落）はテキスト、朗読に拘らず、女性だけが記憶していた。「比較的強い体質をもった私は、」（「上三」「帰りましょう」の段落）「板の隙間から」（「上三」「眼鏡の紛失」の段落）は、男性だけが記憶していた。

また、定性的再現割合を前述の意味単位を利用して、題材の総意味数と1回目に記述したあらすじの意味数から計算した。

【定性的再現割合】

$$= \text{記述意味数} \div \text{題材の総意味数} \times 100$$

題材の総意味数 : 「上三」44 , 「下四十五」42

記述意味数 : 記述したあらすじに含まれる意味数

表4は、上述の再現割合の題材別被験者別の平均である。意味的な観点から見て、被験者は平均して内容の約20～45%記憶しており、「上三」では「見る」の記憶量が多く、「下四十五」では「聞く」の記憶量が多かった。

表 4 題材別被験者別の再現割合の平均

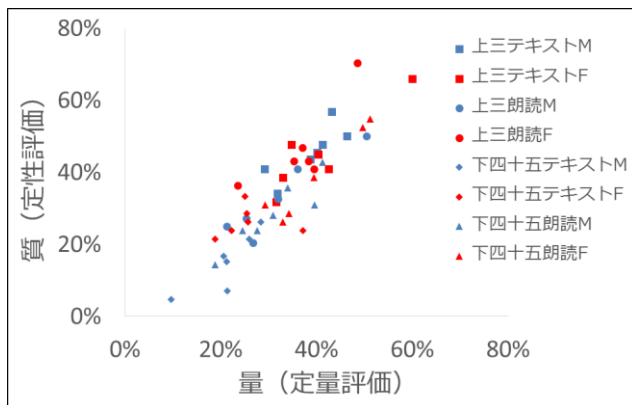
	聞く	見る
上三	40%	45%
下四十五	33%	21%

2.4 定量的な観察×定性的な観察

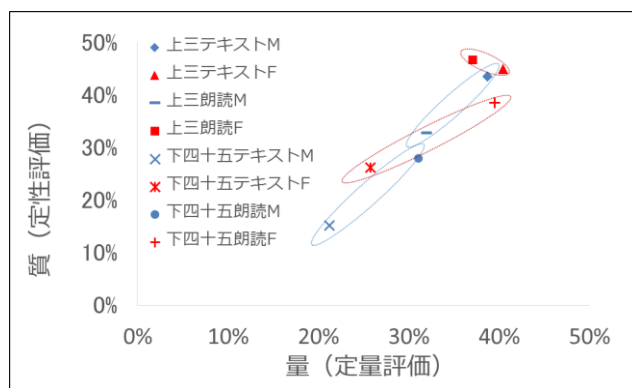
グラフ4に量と質（定性的な再現割合と定量的な再現割合と掛け合わせ）の全被験者の実験結果を散布図として表示する。量（定量的再現割合）は多いが質（定性的な再現割合）が低い、その逆の被験者も散見されたが、全体としては、量が増えれば質も増えるという線形的な相関関係があった（PEARSONの相関係数で0.88）。

グラフ5に素材・被験者ごと（「上三」、テキスト、女性のような条件の組み合わせ）の平均の量と質の散布図（横軸：量、縦軸：質）を示す。量と質の関係が線形的であったため、定性評価で得られた知見に近い結果となった。「上三」では男性は線形的にテキスト優位であり、女性は量ではテキスト優位であるが質では若干朗読優位である。また、「下四十五」では、男性は量質共に線形的に朗読優位であ

るという「上三」と逆の結果が出ている。一方、女性も男性と同様に量質共に線形的に朗読優位であるという結果となった。



グラフ 4 量と質 (全員)



グラフ 5 量と質 (素材・被験者分類ごと)

2.5 視聴覚実験の観察まとめ

「見る」及び「聞く」の違いを定性的、定量的に観察することで得られた知見を以下にまとめておく。

- 約 900 文字、朗読にして約 3 分半程度の文章では、人は大体内容の 20~40% を記憶していた。
- 題材の内容が叙述的であれば「見る」場合の記憶量が多く（「上三」）、感情的であれば「聞く」場合の記憶量が多かった（「下四十五」）。
- 性別で見ると、相対的に女性の方が男性より記述量が多く、「聞く」場合の方がその傾向が強かった。

その他、実験中、あるいは分析の際に気が付いた定性的な知見も記しておく。

- 朗読音声を聞いている時、男性は目を閉じて聞いていたが（参加者全員）、女性は目を閉じずに斜め上を見ている人が多かった。
- 朗読実験の方が自分の言葉で書き起こしていた。テキストより朗読の方が言葉の置き換えが起りやすい。

3. 音響特徴、言語特徴からのアプローチ

人間が「いい加減に」理解している受信内容を何らかの形で保持しておくような仕組みは、様々なシーンで有用なはずである。過去にも多くの要約研究がなされ、特に自然言語処理の世界では、ある程度確立した要約の技術要素がある。音声処理の世界でも、強調箇所や間などの特定技術が利用されることが多く、筆者らが行うアプローチが決して新しいわけではない。ただ、音声と言語の双方からの要約のアプローチを実際の人の記憶と突き合わせて検証するような試みはあまりなされていないのではないと思われる（重要な部分を特定させる、あるいは被験者にテキストを要約させるようなものはあるが、自然に残る記憶をテーマにしたものは稀と思われる）。本稿では、いい加減な受信情報、すなわち人の記憶に残る部分と音響特徴、言語特徴の両側面からの要約結果とどの程度整合するかの検証を行う。

3.1 音響特徴からのアプローチ

大前提として、音響特徴量からみた強調部分には、話者の意図が表れていると仮定している。まず、前述の意味単位ごとに何%の被験者の記憶に残っていたか（以下、記憶度という）を算出し、記憶箇所を比較した。具体的には、「聞く」と「見る」の記憶度の相関係数（PEARSON）は、「上三」で 0.81、「下四十五」0.70 であった。また、定性的に朗読を聞いて強調していると思った箇所を 1、それ以外は 0 として「聞いた」場合の記憶度との相関を確認したところ、「上三」で 0.06、と非常に小さく、「下四十五」で 0.37 であった。また、実験素材ごとの記憶度平均（記憶人数割合）は、誰も記憶していなかった発話区間を除外して計算すると表 5 のようになり、同じ箇所を記憶していた人は平均すると半分もないことが分かる。

表 5 素材ごとの「見る」と「聞く」の記憶度平均

	聞く	見る	総平均
上三	48%	54%	51%
下四十五	39%	29%	34%
全体	43%	42%	43%

「下四十五」は記憶量が少ないにも関わらず、記憶箇所がバラついていた。記憶量・記憶箇所・記憶度平均の比較から、「下四十五」の方が朗読の影響を受けていると考えられる。

次に、音響特徴量からみて強調されている発話区間を抽出した。発話区間と無音区間の分類には、音声分析ソフト Praat(version 5.4.09) (a)を用い、朗読音声を発話区間と無音

a <http://www.fon.hum.uva.nl/praat/>

区間（無音区間の条件は、最小ピッチ：10Hz／無音閾値（dB）：-25／無音時間：0.75 以上／音声最小時間：0.1）に分類した。また、抽出した発話区間と意味単位とで区切りが違ふ箇所については、複数の意味単位と発話区間が紐付く場合は、一番高い記憶度の意味単位を採用し、複数の発話区間と意味単位が紐付く場合は、一番重要な単語を含む発話区間を採用し、前後の「間」は意味単位の区切りで取得するという方法で調整した。また、使用する音響特徴量を決定するために、調整済みの発話区間の音響特徴量を抽出し、最大値・最小値・平均・標準偏差（Pitch／Intensity／HNR／Jitter／Shimmer）に加えて、話速（発話区間のモーラ数÷発話区間の秒数）、発話区間の前後の「間」（無音区間の秒数）を計算した。最終的には、特徴量と発話区間の記憶度の相関から、以下4つを採用した（話速は記憶度と負の相関があり、遅いほど記憶度が高い）。

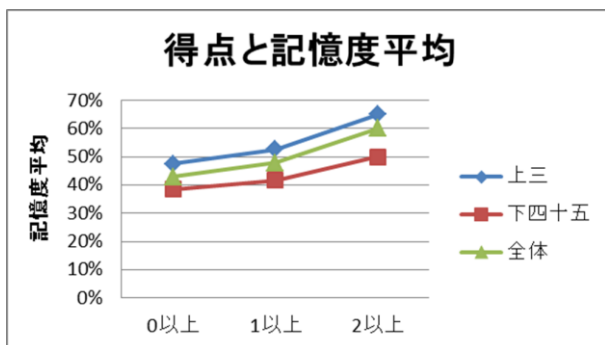
- Pitch の最大値
- Pitch の標準偏差
- Intensity の最大値
- 話速

次に、音響特徴量ごとに音響特徴量の分布から閾値を設定（それぞれ 20%前後が該当するように閾値を設定）し、強調している発話区間を抽出した。

表 6 音響量特徴の強調の閾値

採用した音響特徴量	閾値
pitch_max	>350
pitch_std	>55
intensity_max	>85
話速	≤6

発話区間の得点は、音響特徴量毎に、閾値を超えると1点、それ以外は0点として得点を求め、それらを合算したものとした。以下に、発話区間の得点と朗読の場合の記憶度平均との関係（得点3以上の発話区間が少なかつたため得点2以上までを表示している）を示す。

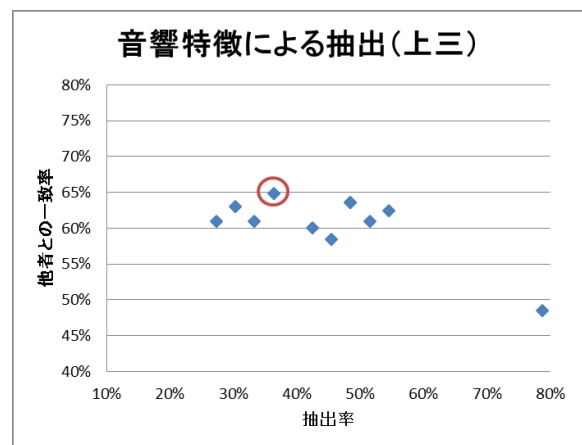


グラフ 6 音響スコアと記憶度平均

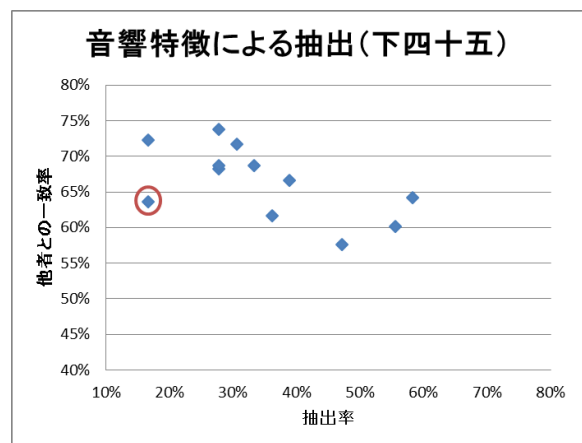
閾値の得点が高いほど記憶度平均も高くなり、音響特徴量を用いて人の記憶に残りやすい文章を選択できてきていると言える。

これを重要箇所抽出システム（システムが重要箇所として抽出する発話区間の得点の閾値を2以上とした）として評価し、その性能を人間と比較した。

評価対象は被験者（「上三」の被験 10 人、「下四十五」被験者 11 人）+システムの 11 人または 12 人であると考えた。一致率（質的指標）は、各々の被験者が記憶していた意味単位ごとに、自分が重要箇所として抽出した（記憶していた）かしないかの選択が他の被験者の内何人と一致していたかであり、抽出率（量的指標）は、全発話区間中何%抽出したかである。他者との一致率を見ると、人間の能力にもバラつきがあり、システムもそのバラつきの範囲内にプロットされている。結果的には、人間と比べても遜色がなかったといえる。上三では、一致率で、すべての人を上回っている。



グラフ 7 音響特徴による抽出 (上三)



グラフ 8 音響特徴による抽出 (下四十五)

以上をまとめると、

- 「上三」に関しては「聞く」と「見る」の記憶度平均に大きな差はなく、音響特徴量からみた重要箇所と、記憶箇所の一致が多くみられた（内容から重要と判断して記憶に残ったのか、朗読の影響なのかは

不明だが、朗読者が重要だと考えた箇所を強調して読んでいたとも推察される.)

- 「下四十五」に関しては、「見る」より「聞く」方が記憶箇所に共通性があり、朗読の影響を受けており、音響特徴量からみると強調している箇所が少なかった。
- 「重要箇所抽出」に関しては、利用した音響特徴量も少なく、システムの仕様としても単純であったにも関わらず、人間と比較しても遜色のない結果が得られたことから、音響特徴量、音声の非言語情報の有効性が認められた。但し、実際に運用する場合には、特定話者向けのチューニング、対話や講演など発話スタイルや場面による違い、「間」などのその他知見を考慮するなど課題は数多くある。

その他、気が付いた点を記しておく。

- 朗読の影響を受けていると思われる「差し上げるなんて威張った口の利ける境遇ではありません。」という箇所は、「見る」では誰の記憶にも残っていなかったが、「聞く」では73%の人の記憶に残っていた。
- 朗読の影響ではなく内容から重要と判断されたとと思われる「お嬢さんを下さい」という箇所は、結婚を申し込む重要なセリフであるが、抑えた表現で読まれており、音響的な特徴は出ていなかったが、ほぼ全員の記憶に残っていた。「下四十五」は音響特徴量からみると強調している箇所が少なかったが、感情的な内容で、読み方を抑えても十分内容が伝わるものだった。あえて抑えて読む方がかえって記憶に残るという朗読者のテクニックなのかもしれない。

3.2 言語特徴の利用

音響特徴が朗読音声による記憶と密接な関係があることが示された。それでは、言語特徴を利用して抽出したものは、人の記憶とどのような関係があるのだろうか。また、音響特徴と言語特徴を組み合わせることで、人の記憶に残りやすい重要箇所を抽出することができないのだろうか。それらを確認するために、言語特徴を利用した重要箇所抽出を試み、その結果と人の記憶に残っていた部分の突合せ評価を行い、さらに、音響特徴と言語特徴を利用して重要箇所抽出を試み、その結果と人の記憶に残っていた部分の突合せ評価を行った。音響特徴で行った実験との比較のために評価する意味単位は、音響特徴による評価実験の単位に合わせた。

言語特徴からの重要箇所の特定には、次に示すような過去から多く試みられているオーソドックスな方法を選択した。まず、テキストの定量的な特徴を知るために、青空文庫の小説の十分な量のテキスト (16MB) と実験対象の 2

つのテキスト (どちらも 2K ほど) を混ぜて、実験対象のテキストの形態素数に合わせて分割したドキュメントを元にして実験対象テキスト及び被験者の書き起こしテキストから出現単語の TF-IDF 値を求めた。意味単位の重要指標に関する言語特徴スコアはその意味単位に含まれる TF-IDF 値の合計を意味単位のスコアの平方根で割ったものを採用した。求めた言語特徴を重要箇所抽出のために適当な閾値で分割した (付与された言語特徴スコアの分布を見て、9.5 以上を重要箇所として抽出した)。

また、前項で述べた音響スコアと定義した言語スコアを定義する都合上、求めた言語スコアを分布から見て適当な閾値で 0-4 の値に離散化し、それを言語特徴スコアとした。音響+言語スコアは、表 7 に示すテキスト記憶度、朗読記憶度と音響、言語それぞれの相関係数 (PEARSON) の平均を重み係数 (音響特徴スコアの重み係数が 0.35、言語特徴の重み係数が 0.21) として各スコアの加重平均を求めたものとした。表 7 に各特徴のスコアと被験者のテキスト記憶度、朗読記憶度との相関を示す。記憶度は、前述のとおり、被験者のうち何%の人の記憶に残っていたかを表す数値指標である。表 8 には、各特徴によって抽出された重要箇所と被験者のテキスト記憶度、朗読記憶度との相関を示す。

表 7 各特徴とテキスト記憶度、朗読記憶度との相関

相関係数 (Pearson)		音響特徴		言語特徴		音声+言語特徴	
		朗読	テキスト	朗読	テキスト	朗読	テキスト
特徴スコア	上三	0.46	0.41	0.10	0.25	0.44	0.45
	下四十五	0.32	0.22	0.24	0.23	0.41	0.34

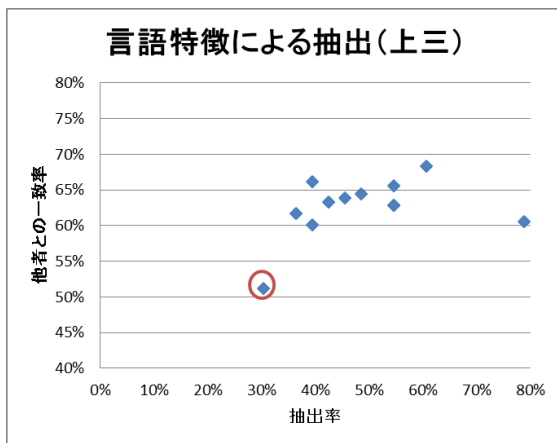
表 8 抽出結果とテキスト記憶度、朗読記憶度との相関

相関係数 (Pearson)		音響特徴		言語特徴		音声+言語特徴	
		朗読	テキスト	朗読	テキスト	朗読	テキスト
重要箇所抽出	上三	0.58	0.55	0.05	0.25	0.58	0.55
	下四十五	0.32	0.21	0.24	0.23	0.35	0.26

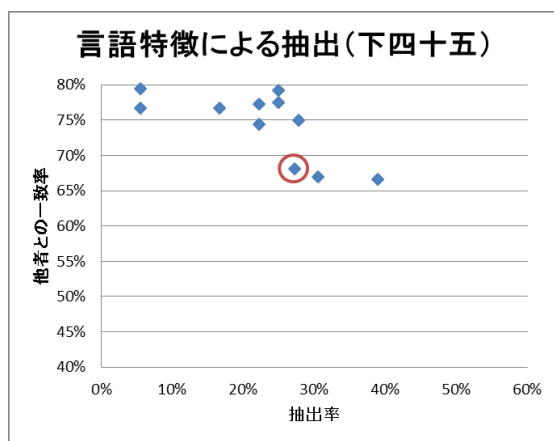
両方の表を見ると、このテーマにおいては、音響特徴の方が言語特徴よりも重要箇所の抽出課題に強いといえる。音響特徴+言語特徴では単体での評価より若干改善がみられる (特に下四十五)。また、音響特徴は朗読からの抽出課題に優位性が見られるが、朗読からの抽出においてもテキストからの抽出においても、有効そうである。一方、言語特徴は「上三」においてのみテキストからの抽出課題に優位性がみられる。「上三」における朗読記憶度と言語特徴はほぼ無相関である。

グラフ 9,10 に音響特徴からの重要箇所評価と同様の評価 (重要箇所抽出システムとして評価し、その性能を人間と比較、評価対象は被験者「上三」の被験 10 人または「下四十五」被験者 11 人+システムの 11 人又は 12 人であると考え、一致率は、各々の被験者が記憶していた意味単位ごとに、自分が重要箇所として抽出したかしないかの選択が他の被験者の内何人と一致していたかであり、抽出率は、

全発話区間中いくつか抽出したかである)を行った。



グラフ 9 言語特徴による抽出 (上三)



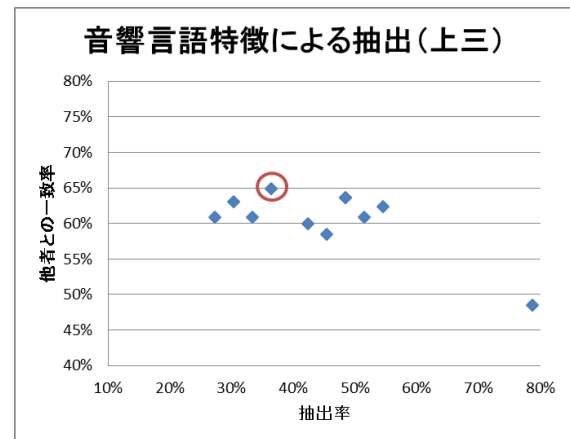
グラフ 10 言語特徴による抽出 (下四十五)

グラフ 9,10 を見るかぎり, 言語特徴による抽出システムは人の記憶との関連が弱くうまく抽出できているとはいえない。「下四十五」では, 一致率の低い被験者より一致率が高かったが, 「上三」ではすべての被験者の一致率を下回った。

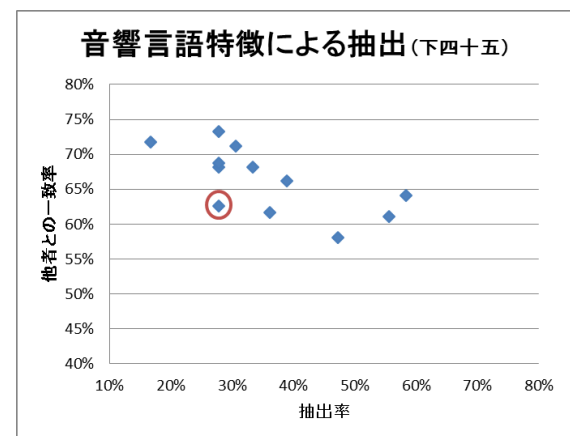
グラフ 11,12 は音響特徴+言語特徴による抽出システムの評価である。「上三」では, 音響特徴のスコアが効いており, 言語特徴のスコアが反映されるものはなく, 音響特徴スコアで評価した時と同じ評価結果になった。然しながら, 「下四十五」では, 言語特徴スコアが反映されて音響特徴スコアだけの時よりも若干抽出率が大きくなっている。抽出率が大きくなっても一致率は低下しなかった。言語スコアで抽出した内容を見ると, 音響特徴スコアだけでは抽出できなかったが, ほぼ全員の記憶に残っていた「下四十五」の「お嬢さんを下さい」「下さい, ぜひ下さい」という結婚を申し込む重要なセリフが, 言語スコアを加えることで抽出されていた。このようなことを考慮すると内容に依存する部分は言語特徴の得意部分で, 話し手の話し方に依存す

る部分は音響特徴からの得意部分というように役割分担することで人の記憶に残りやすい部分の抽出が可能になるかもしれない。

今回の実験の題材は, 1 世紀前の小説であり普段話されている言葉とはかなりの乖離がある。また, 朗読の読み手はプロである。そのあたりの影響がどの程度出ているのかも気になるところである。



グラフ 11 音響言語特徴による抽出 (上三)



グラフ 12 音響言語特徴による抽出 (下四十五)

4. おわりに

言語というインターフェースを介したコミュニケーションの「いい加減さ」に着目し, 小説のテキストを見るという行為と朗読を聴くという行為で得られる情報を人がどのように解釈し, それぞれの入手方法によってどのような違いがあるのかを知るための実験を紹介し, 言語特徴や音響特徴から抽出した箇所と人の記憶に残ったものとを比較することで受信側の「いい加減な」コミュニケーションを観察した。そして, 音響特徴と言語特徴の役割を知ることによって人の記憶に近い要約の実現可能性を示した。

次の目標として, 本稿で紹介した「いい加減な」コミュ

ニケーションの仕組みを対話システムの中に組み込むことを検討している。アイデアとしては、時間軸のパラメータを考慮したロボットとユーザー発話の要約と生成する発話の関係性モデルを構築することで発話生成を支援するような仕組みである。具体的な実現方法は決まっていないが、今後そのような取り組みの中で今回の知見を活かしていきたいと考えている。

参考文献

- 1) マイケル・S・ガザニガ：人間らしさとは何か？，インターシフト，(2010).
- 2) マイケル・S・ガザニガ：脳の中の倫理，紀伊国屋書店，(2006)
- 3) 谷田泰郎，高椋琴美，津田沙織：いい加減な対話からの心のモデルの抽出，人工知能学会全国大会（第29回）JSAI2015，(2015)
- 4) 高椋琴美，谷田泰郎：「聞く」と「見る」における言語理解の違い，日本音響学会2015年秋季研究発表会，(2015)
- 5) 樋渡涓二：視覚と聴覚はどう違うか，テレビジョン 31(11)，P853-861，(1977)
- 6) 滝田亘，中山 実：視覚と聴覚による文章の提示と記憶への影響，日本教育工学雑誌 27，P81-84，(2004)
- 7) 濱田治良：短期記憶における視覚記憶と聴覚記憶の差異，心理学研究 61，8-14，(1990)