

いい加減な対話からの心のモデルの抽出 — 「聞く」と「見る」における言語理解の段階的実験 — Extraction of models of mind from sloppy communication

谷田 泰郎*¹ 高椋 琴美*¹ 津田 沙織*¹
Yasuo Tanida Kotomi Takamuku Saori Tsuda

*¹ シナジーマーケティング株式会社
Synergy Marketing, Inc.

Improving marketing practices supposes a better understanding of how the human mind is designed and how it interacts with others. We focus in this paper on the sloppiness of human communication. We compare experimentally the difference in recollection of a text when it was communicated in an oral or a written form.

1. はじめに

筆者らは、心理的精神の便益こそが最も基本的なニーズであり、マーケターが実現できる究極の差別化であるという考え[1]に基づき、人と人、人と物の出会いから生まれるコトに対して感じる価値をより大きなものにするために、「心のモデル」を介在させることで「感じる」をより心地よいものにするという価値観マーケティング戦略を取っている。そのためには、デジタル情報空間に存在する行動データから効率的に価値創出を行う必要がある。具体的には、ビジネスモデル化された研究成果として、個性としての価値観とその成分により説明される社会的類型である *Societas* を利用して、様々な関係性モデルを構築することで、巨大なデジタル情報空間内での効果的なコミュニケーションを実現しようとしている[2]。

ヒトは大量の刺激情報を入力として受け取り、脳の中で高度な情報処理演算を行なって、行動を出力する。ヒトが受信した入力情報とヒトが発信した出力情報の関係性をモデル化するだけでは、ヒトの行動理由までは見えてこない。ヒトとヒト、ヒトとモノとの出会いがコトを産出するのだとすれば、コトとは、情報との出会いの際に心の中で起こっている変化だと考えられる。然しながら、現状の理性的な枠組みだけで、その変化を適確にとらえることはできない。遺伝と環境、社会との関係性、中長期的な適応に対する変化、感情や情動といった短期的な変化を整理していく必要がある。

我々の研究グループは、コミュニケーションの中に「心のモデル」や意識を解くためのエビデンスがあると考えている。ガザニガの著書[3][4][5]によれば、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。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なのだ。生物は生得的に生きて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。学ば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。筆者らは、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と考えている。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。私たちは常に合理性と直観の中で戦っており、どちらが勝利しても自分自身の矛盾を矛盾がないものとして解釈する必要があるのだ。

このような「いい加減な」コミュニケーションを通じてより心地良さを感じる「心のモデル」をデザインするために、特定の人々の発話を学習して発話を生成する個性的なキャラクター（ロボット）と対話できるゲーム感覚のアプリをリリースすることで対話データの収集を行う計画（図1）もあり、対話ロボットに感情移入できるかの仮指標として、ロボットに個性を感じることができているのかを確かめるための実験も行っている[6]。

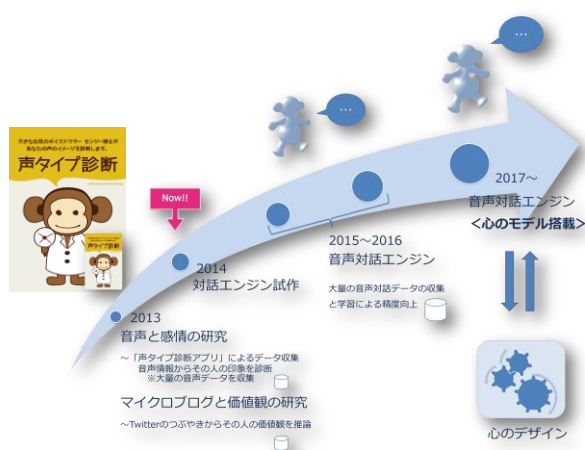


図1 音声対話データからの心のモデルの抽出

本稿では、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、手始めに、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

連絡先：谷田泰郎，シナジーマーケティング株式会社
研究開発グループ，電話番号：06-4797-2909，メールアドレス：tanida.yasuo@synergy101.jp

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ので、その報告をする。

2. 視覚と聴覚に関する一般的な知見

視覚も聴覚も触覚など他の感覚と同様に視床を経由して大脳皮質に伝えられることから、他の感覚モダリティの影響を受ける。視覚と聴覚が同時に働いたときに起こる視覚優位の例として、口元の動きと聴いている音声が違うとき錯聴が起きるといったマーカー効果[7]などがよく知られているが、聴覚の優位な特徴としては、360度どの方向からでも情報を入力できるという点と時間特性の認識が挙げられるだろう。神経脳活動のような細かい時間的特徴の識別を要する場面では、オシロコプの画像より波形を音にした方がよいというわけだ。視覚刺激の呈示頻度が一定でも、聴覚刺激の呈示頻度が変化すると視覚刺激の呈示頻度の判断が変わってしまうというように、時間的特徴の識別場面では聴覚情報が優先される聴覚優位が起きるといえる[8]。

視覚情報は網膜から入力され、中継核を介して視床に到達し、その後視覚野(V1)に送られる。V1の単純神経細胞は特定の位置・特定の傾きをもつ光線刺激に反応し、その刺激選択性の反応パターンは2次元のGabor関数で近似できることが知られている[9]。ピクセル単位のナイーブな画像表現を神経細胞のコストをできるだけ小さく情報をできるだけ落とささないようにするために、V1の表現は小さな線分を基本とする最適な基底となっている。スパースコーディング仮説によれば、少ないビット数で情報を落とさずに(最優先でエッジを表現すると効率がよい)最適な表現を与える基底を探索する最適化問題を解いたところ、その基底はGabor基底とよく似ていたという[10]。つまり、V1で見られる刺激選択性は自然画像を入力とした最適化問題を解いた結果(自然画像に対する適応の結果)なのだろう。

一方、聴覚情報は蝸牛から入力され、視覚情報と同様に中継核を介して視床に到達し、その後聴覚野(A1)に送られる。A1の解明はV1のように進んでいない。A1の細胞の選択性は多様である(200Hzのみに反応、400Hzに反応、どちらにも反応するなど)。自然画像には局所的な特徴が多く特徴と視野角の相関が点でないが(距離が大きくなるほど特徴差異が大きくなる)、自然音の場合、近い周波数には局所的な相関があるものの、選択性が多様なために(離れた周波数にも強い相関がある)非局所的な統計性(A1は周波数軸上で分散した基底を持っている)がある。このように、自然統計性が複雑なために学習結果である神経生理学的特徴が複雑で別物に見えるが、V1とA1の学習戦略は別物ではなく自然統計性の差異を反映しているのだという[11]。

過去に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[12][13][14]も数多くあるが、決定的な結論には至っていない。視覚と聴覚に限らず(筆者らはそこに絞っているが)感覚モダリティを解き明かすためには、言語というもう一つの普遍的な基底を抽出する必要があるのではないだろうか。

3. 視聴覚実験の概要

私たちは、視覚と聴覚の様々な違いを知っているが、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をあまり知らない。本稿では、それを調べるために、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おうという簡単な視聴覚実験を行った。

具体的には、夏目漱石の小説「こころ」(「上三」の章の前半4段落及び「下四十五」の章の前半3段落)の朗読音声を「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な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ことで、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: 実験の条件など

	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
			1回目	2回目		3回目
午前	11人	上三	11.5分	15分	下四十五	11分
			15分	18分		15分
			18分	17分		17分
午後	10人	下四十五	11分	15分	上三	11分
			15分	18.5分		15分
			18.5分	18分		18分

その他、実験終了後に、「日常的に読書をしますか?マンガや雑誌も含め、どのような本をいつ・どのくらい読むかを教えてください。」「読むこと・聞くことに関して、自分に得意不得意があると感じますか?」「本日の調査を振り返って、自分の回答に関するコメントや進め方・内容・分量・時間・その他全般への感想などご自由にお書きください。」という3つのアンケートにフリーアンサーで答えてもらった。その他、被験者の年齢、性別、3種類の価値観(Societas, Schwartz, big5)を聴取した。

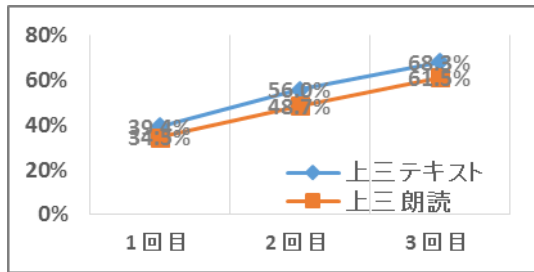
4. 視聴覚実験の結果と考察

人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。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)

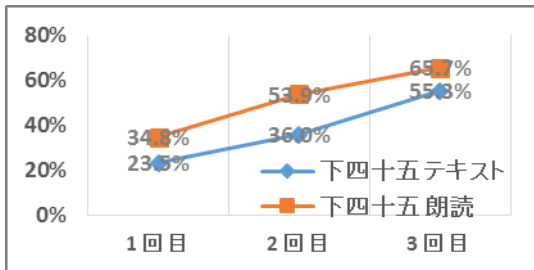
に応じて計算コストのかからない方法で重要な部分を抽出しようとしているはずで、その様子を観察することはできる。そこで、テキストを入力にした場合と朗読を入力にした場合でどのような違いがあるのか、重要であるものを抽出しているかどうか、男女の違いはあるのか、などに焦点を絞り、定量的、定性的な側面から評価した。

4.1 定量的な観察

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。正確には間違っているものも含まれるが、この数字を量的な記憶の再現割合の指標とみなす。



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

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 비해、「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）。「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述的な記述が多い。一方、「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。このように、定量的な再現割合では、対象の内容が叙述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。

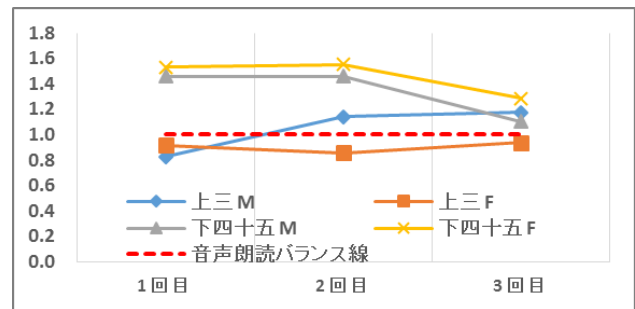
表2は各実験条件における男女別のテキスト及び朗読に対する記述量（再現割合）の変化である。男女別では、全体的に女性の方が男性より記述量が多い。テキストと朗読

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になる。



グラフ3: 朗読再現割合/テキスト再現割合

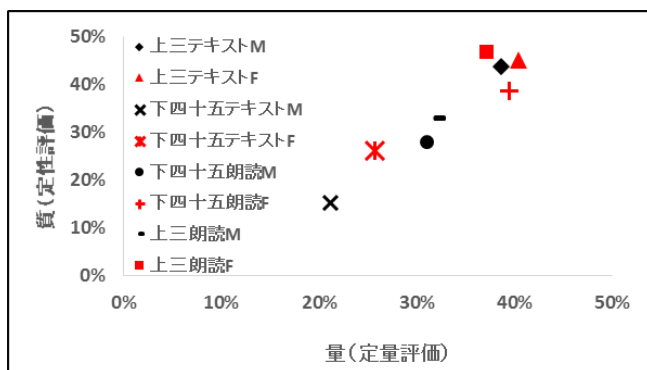
その他、テキストの定量的な特徴との関係を知るために、青空文庫の小説の十分な量のテキスト（16MB）と実験対象の2つのテキスト（どちらも2Kほど）を混ぜて後者の形態素数に合わせて分割したドキュメントを元にして実験対象テキスト及び被験者の書き起こしテキストから出現単語のTF-IDF値を求め、上位30語を比較した。被験者の書き起こしテキストでは同じ意味での置き換えが起こっているため目検で一致数を調べた。紙面の関係上詳細な報告はしないが、後述の定性評価などで成績の良かった被験者が機械的な重要語の抽出率が高く、朗読よりもテキストを入力にした時の方が機械的に抽出した重要語との一致率が高いという結果で、当初から予想された範囲内に収まった。

4.2 定性的な観察

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く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。本稿では煩雑過ぎるため表示しないが、定性的な再現割合を前

項で述べた定量的な再現割合と掛け合わせて全被験者の実験結果を散布図として表示すると、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 4 に各実験（「上三」、テキスト、女性のような条件の組み合わせ）における平均の質の散布図（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る。一方、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：定量的再現割合と定性的再現割合

また、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つたでしょう。”（「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」の段落のように叙事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。

5. おわりに

朗読実験の時、男性は目を閉じて音声を聞いていたが、女性は目を閉じずに斜め上を見ている人が多かった。朗読とテキストの違い、男女の差について述べてきたが、本稿の定性的な評価の中では多くは同様の評価であって、朗読でもテキストでも、男性でも女性でも反応する箇所は似たようなものなのである（その評価結果を PEARSON の相関係数で表すと朗読・音声で 0.77、男女で 0.72 であった）。ただ、その小さく見える違いが実は大きな違いなのである。

「いい加減」なコミュニケーションを解明していくために、小説のテキストと朗読による簡単な視聴覚実験を行ったが、多くは未解明である。本稿で紹介したデータは、まだ手つかずの部分が多い。例えば音声に関する特徴（例えば音響特徴量と記憶に残る部分の関係）について調べていないし、価値観やアンケートのフリーアンサー部分についても精査していない。入力言語をどのように「いい加減」に解釈して伝えるのか、どのように感じるのか、その時どんな個人差があるのか等を理解してより良いコミュニケーションを構築するためにも、まずは既存のデータ分析と未知のデータ収集の仕組みを構築していく予定である。

参考文献

- [1] P.コトラー, H.カルタジャヤ, I.セティアワン: コトラーのマーケティング 3.0, 朝日新聞出版, (2010)
- [2] 谷田泰郎: 価値観マーケティングと社会知ネットワーク, 人工知能 9月号 Vol.29, No.5, P456-463, (2014)
- [3] マイケル・S・ガザニガ: 私はどこにあるのか?, 紀伊國屋書店, (2014)
- [4] 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)
- [5] マイケル・S・ガザニガ: 脳の中の倫理, 紀伊國屋書店, (2006)
- [6] 谷田泰郎: グローバル・コミュニケーションと個性, 電子情報通信学会技術研究報告 Vol.114 No.465 P19-21, (2015)
- [7] McGurk Harry, MacDonald John: Hearing lips and seeing voices, Nature 264(5588), 746-748, (1976)
- [8] Recanzone, G. H: Auditory Influences on Visual Temporal Rate Perception, Journal of Neurophysiology 89, 1078-1093, (2003)
- [9] D.H. Hubel and T.N. Wiesel: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, The Journal of Physiology, 160, 1, pp.106-154, (1962)
- [10] B.A. Olshausen and D.J. Field: Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature, 381, 6583, pp.607-609 (1996)
- [11] 寺島裕貴: 脳の画像・音声処理戦略を解き明かすスパースモデリング, 映像情報メディア学会誌 Vol. 68, No. 12, P897~901, (2014)
- [12] 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)
- [13] 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)
- [14] 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)