

# 対話コミュニケーションにおける 身体動作の同調と発話内容との関係性の分析

○三浦 俊介 (東京工業大学), 横塚 崇弘 (東京工業大学), 権 眞煥 (東京工業大学),  
Thepsoonthorn CHIDCHANOK (東京工業大学), Yap MIAO SIN ROBIN (東京工業大学),  
小川 健一朗 (東京工業大学), 三宅 美博 (東京工業大学)

## Analysis of the relationship between bodymotion synchrony and speech content in human communication.

○Shunsuke Miura (Tokyo Institute of Technology), Takahiro Yokozuka (Tokyo Institute of Technology), Kwon Jinhwan (Tokyo Institute of Technology), Thepsoonthorn Chidchanok (Tokyo Institute of Technology), Yap Miao Sin Robin (Tokyo Institute of Technology), Ken-ichiro Ogawa (Tokyo Institute of Technology), and Yoshihiro Miyake (Tokyo Institute of Technology)

Abstract: Human communication has a couple of channels, verbal channel and nonverbal channel. Many previous studies have focused on each channel. However, the relationship between the two channels has not been yet revealed. Therefore, the purpose of this study is to analyze the relationship between verbal channel and nonverbal channel in face-to-face communication by focusing on the speech content and bodymotion synchrony.

**Keywords:** face to face communication, speech content, bodymotion synchrony

### 1. はじめに

人間の対話コミュニケーションには、発話の意味内容といった言語チャンネルと身体動作といった非言語チャンネルがある。

言語チャンネルに関する研究として、会話における流れや特徴語に着目した研究がある。例えば、課題解決を目的とした会話の流れが組織の中での上下関係によって異なることが報告されている<sup>[1]</sup>。また、会話文における「嬉しい」などの感情を示す特徴語や文型パターンに注目した感情推定手法が提案されている<sup>[2]</sup>。さらに、製品の設計についての会話を出現頻度の高いキーワードを用いて分析を行った研究がある<sup>[3]</sup>。

一方、非言語チャンネルに関する研究として、身体動作の同調に着目した研究がある。例えば、話し手が聞き手に情報を与える一方向型の対話コミュニケーションにおいて、聞き手の理解度が高い場面で二者の身体動作が同調する傾向を確認した研究がある<sup>[4]</sup>。また、二人が話し合っただけの対話コミュニケーションにおいても、二者間の合意度と身体動作の同調との間に関係性がみられた<sup>[5]</sup>。

このように、これまで言語チャンネルと非言語チャンネルそれぞれのチャンネルに注目した研究が多くされてきた。しかしながら、両者の関係性についてはまだ定量的に明らかにされていない。

そこで本研究では、非言語チャンネルである身体動作

の同調と言語チャンネルである発話内容との間にある関係性を分析する。研究の方針として、井上らが行った教師課題による実験のデータ<sup>[6]</sup>をもとに分析を行う。言語チャンネルに関しては音声データから発話内容を分析し、非言語チャンネルに関しては頭部につけた加速度センサのデータから身体動作の同調を分析し、それぞれの分析結果から両者の関係性を調べる。

### 2. 使用する対話データ

本研究では、井上の研究で行われた一方向型の対話実験におけるデータを使用する<sup>[6]</sup>。以下にその実験の詳細を記載する。分析に用いたデータセットは女性2名(1ペア, 23歳)の音声データおよび加速度データである。

#### 2.1 実験課題

この実験は課題として教師課題を設定している。教師課題とは二人が対面して座っている環境で、教師役となった参加者が生徒役となった人に向かって情報を伝達する課題である。

課題で用いられた教師役への資料は、「先人が学問的に考えた人生の意義」について書かれたものである。これは Wikipedia の記事ページ「人生の意義」から引用されたもので、二つのパートに分けて作成された<sup>[6]</sup>。一つ目のパートは共感が得られやすいように専門用語や人物名などを極力省き、参加者にとって理解しやすい共感的内容である。その逆に、二つ目のパートは造語や人名などが含まれた複雑で理解しにくい伝達の内容である。

## 2.2 実験環境

実験は、対話に適した室温、湿度、明るさの個室にて行った。二人の実験参加者は約 1.5[m]の距離をあけてテーブルを挟んで向かい合わせに着席した。テーブルの上にはふたつの書見台がそれぞれ参加者から見やすいように置かれ、そこに課題資料を固定した。身体運動の取得のためにサンプリング周波数 100[Hz] の三軸加速度センサ(小型無線ハイブリッドセンサ WAA-006, ワイヤレステクノロジー社製, 日本)を準備し, 参加者それぞれの額にゴムバンドで固定した(Fig.1)。また, 音声データの取得のために片耳式ヘッドセット(CC530, SENNHEISER COMMUNICATIONS 社製, ドイツ)を装着し, 録音を行った。実際の実験環境を Fig.2 に示す。

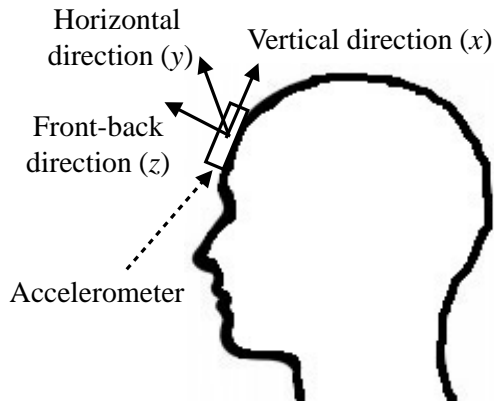


Fig.1 Equipment of the acceleration sensor.



Fig.2 Experiment environment.

## 2.3 実験手順

はじめに実験者は参加者に実験内容とルールを説明し, 参加への同意を得た。その後, 生徒役は退出し, 実験者は教師役に資料を渡して熟読と生徒役への説明の練習を行うように指示した。このとき生徒役として実験者は教師役の向かいに着席し, 練習が十分にされたかを主観的に判断した。練習が終了したら, 生徒役を呼び出して実験装置の取り付けを行った。準備ができたあと, 実験者は対話実験開始の合図を行い, 退出した。実験室には

参加者二人だけが残し, 教師役が生徒役に与えられた内容について説明をした。教師役は説明が終わり次第, 呼び鈴を鳴らして実験者を呼び出した。実験者は対話実験の終了を合図して計測を終了した。その後, 教師役は退室し, 実験者は生徒役にアンケートについて説明を行った。生徒役はアンケートに答え, 教示された内容理解のテストを実施した。制限時間はなく, 生徒役が納得のいくまで解答を行ってもらい, 終了したあとにその場で答え合わせを行った。そして教師課題による対話実験は終了となった。

## 3. 発話内容の分析手法

発話内容分析は坂原の提案手法を参考に行った<sup>7)</sup>。はじめに, 実験で得られた音声データをテキストへと変換する。整形されたテキストデータをもとに形態素解析して単語を抽出する。そして, 発言一回を一区間として, 全区間から単語間の類似度を算出する。算出された類似度をもとに単語クラスタリングを行い, 一発言ごとの特徴ベクトルを作成する。抽出されたベクトルを用いて隣接する発言の変化量を計算し, シーンの分割箇所を特定して分割をする。この単語クラスタリングからシーン分割までの流れを再帰的に行うことで, 話し手の全発言をシーンという少ないまとまりに分けて見ることができる。

### 3.1 テキスト化

本研究では発話内容を分析するにあたり, 話し手の発言が録音された音声データから発話内容をテキストへと書き起こす。まず, 音声認識ソフトウェア(AmiVoice SP2, 株式会社アドバンスト・メディア製, 日本)を用いて, 音声データから音声認識によるテキスト化を行う。その後, 音声を聞きながら人の手によって修正を行うことでテキストデータを作成した。

### 3.2 形態素解析

発話内容を単語ごとに分割するため形態素解析を行う。形態素解析エンジン MeCab<sup>8-9)</sup>を使い, 発言ごとにテキストデータを解析して品詞情報が付与された単語を出力する。ここで, 形態素解析エンジン用に使用する辞書には既存の IPA 辞書<sup>9)</sup>を用いる。

品詞情報が付与されることを利用し, 単語クラスタリングの際に対象とする品詞の単語を抽出する。坂原らの報告<sup>7)</sup>に習い, 名詞(一般, 固有名詞, サ変接続), 動詞(自立)のみを用いる。

### 3.3 単語クラスタリング

はじめにクラスタリングを行うために, 単語間の類似度を計算する。単語間の類似度として Jaccard 係数:

$$\text{Jaccard coefficient} = \frac{X \cap Y}{X \cup Y} \quad (1)$$

を用いる。ここで,  $X \cap Y$  は二つの異なる単語  $w_i, w_j$  が全発言中で同じ発言に出現した回数を表し,  $X \cup Y$  は単語  $w_i, w_j$  のいずれかが出現した回数を表す。そのため

Jaccard 係数は値が大きいほど類似度が高いことを示す。

クラスタリング手法には初期値に依存されない Affinity Propagation<sup>[10]</sup>を用い、Jaccard 係数による単語間の類似性から単語の集まりであるクラスタを抽出する。

### 3.4 シーン分割

まずシーンの変化点を見る際、窓を用いて検出する必要がある。なぜなら一発言ずつ見ていくとちょっとした言い間違いによる「あ、間違えた」などの感嘆がノイズになり、そこで話題が途切れる要因になりえるからである。そこで発言を複数にまとめて扱う。例えば、窓幅を3としたとき、{1,2,3} と {4,5,6} のまとまりを隣接する窓とする。

次に隣接する窓間の距離を計算する。ある窓における複数発言内の単語が各クラスタにいくつ含まれているかをベクトルとして算出する。そのベクトルを使って隣接する窓 $s_i$ ,  $s_{i+1}$ の距離を次のコサイン距離で導く、

$$\text{Cosine distance} = 1 - \frac{s_i \cdot s_{i+1}}{|s_i||s_{i+1}|} \quad (2)$$

さらに隣接する発言間の距離に重みを掛け合わせる。発言間だけでなくクラスタ同士の距離を重みとして考慮することで、関係性のある話題から話題への遷移における変化を小さく捉えることができる。隣接する窓の複数発言 $s_i$ ,  $s_{i+1}$ の各クラスタに含まれる単語数が $\{C_{i1}, C_{i2}, \dots, C_{in}\}, \{C_{(i+1)1}, C_{(i+1)2}, \dots, C_{(i+1)n}\}$ であるとき、重み $\omega$ を次式で計算する、

$$\omega_{i,i+1} = \sum_{j,k=1}^n \frac{\text{transition}_{i,i+1}(j,k)}{N} \quad (3)$$

$$\text{transition}_{i,i+1}(j,k) = \begin{cases} 0 & (|C_{ij}| = 0 \text{ or } |C_{(i+1)k}| = 0) \\ 1 - C_{ij} \text{ and } C_{(i+1)k} \text{ の代表点の距離} & \end{cases}$$

ここで、 $N$ は $s_i$ と $s_{i+1}$ 中の単語が含まれているクラスタの数を乗算した値である。

次に、重み付けをしたコサイン距離からシーンの分割箇所を判断する。閾値 $\theta$ 以上の最も大きな値をもつ隣接する発言の間をシーンの句切れとして、前後に分割する。その後も分割されたシーンを単語クラスタリングからシーン分割まで再帰的に行い、話し手の発話内容を複数のシーンへと分割する。

## 4. 身体動作の同調の分析手法

身体動作の同調の分析方法として井上の報告を参考に分析する<sup>[9]</sup>。はじめに身体動作の加速度データから合成ノルムを計算する。その加速度ノルムから短時間 Fourier 変換を用いて振幅スペクトルを抽出し、このスペクトルの絶対値を身体動作の指標とする。ペアである二人の指標を用意し、窓を用いて相関分析を行う。ある点における窓が有意な正の相関であり、また、強いうなずきである場合にその点を同調区間とした。

### 4.1 加速度ノルムの計算

本研究では、身体動作として額に取り付けた加速度セ

ンサから取得した頭部運動のデータを用いる(Fig.1 参照)。三軸方向 $x, y, z$ の加速度 $a_x(t)$ ,  $a_y(t)$ ,  $a_z(t)$ から次式を用いて合成ノルムを算出する、

$$a(t) = \sqrt{a_x^2(t) + a_y^2(t) + a_z^2(t)} \quad (4)$$

### 4.2 短時間 Fourier 変換

算出された加速度の合成ノルムをもとに短時間 Fourier 変換：

$$F(v, t) = \int_{-\infty}^{\infty} a(t') \omega(t' - t) \exp(-2\pi i v t') dt' \quad (5)$$

を行う。ここで、 $v$ は周波数であり、 $t$ は窓関数における基準点である。このように、短時間 Fourier 変換を採用することで時間要素を残せるので、周波数×振幅スペクトル×時間という三次元データを扱うことができる。窓関数には Humming 窓 $\omega(t)$ を適用し、窓幅を 1.28[sec] (データ点: 128 点)に設定した。また、窓の移動幅を 0.1[sec]として計算した。

なお、短時間 Fourier 変換で得られる周波数帯にはさまざまな身体動作の振動が現れている。そこで、うなずきの周波数帯付近とされる 1.0 ~ 5.0[Hz]の周波数帯を 0.5[Hz]ごとに調べる<sup>[11]</sup>。さらに、短時間 Fourier 変換の結果の周波数を軸に線形補間を行う。そして、振幅スペクトルの絶対値を算出し、これを身体動作の指標とした。

### 4.3 同調検出

二者対面実験でペアだった二人分の身体動作の指標を計算し、両者間の身体動作の同調を検出する。身体動作の同調を見るために相互相関的手法を用いた。相関分析には Spearman の順位相関係数を用いる。窓幅は 1.8[sec]に設定し、その窓を 0.1[sec]ずつずらしていく。さらに、二つの窓の時間ずれを -0.5 ~ 0.5[sec]までずらしながら相関分析を行う。このとき、時間ずれが正の場合は聞き手が先行しており、負の場合は話し手が先行している。

相関分析の結果を用いて以下の二つの条件を満たす窓の基準点を身体動作の同調区間とした。一つは二者間の相関が有意な正の値であること。もう一つは見ている窓において両者ともに身体動作が発生していることである。一人ずつ窓ごとに身体動作の指標の平均値を計算し、値がその実験全体の指標の 90 パーセント値以上である場合身体動作が発生したとする。

## 5. 結果

### 5.1 発話内容の分析結果

はじめに、音声データをテキスト化した結果を示す。教師役に渡される資料には 24 文の文章が記載されているが、教師役が話した内容は 27 文と資料より多く書き起こすことができた。これは教師役が生徒役に説明をする際に、言い間違いや同じ一文を繰り返して話していることがあったためである。

次に、シーン分割によって得られたシーン出力の結果を表 1 に示す。窓幅を 2 に、 $\theta$ を 0.1 と設定したとき、1、

2 ~ 4, 5 ~ 7, 8, 9, 10 ~ 12, 13 ~ 15, 16, 17 ~ 19, 20 ~ 23, 24, 25 ~ 27 発言目が複数発言のまとまりであるシーンとして出力された。全部で12のシーンに分割され、4文以下のまとまりで構成された。

ここで、シーン分割の際に行われる単語クラスタリングの結果の例を示す。全発言を対象にした場合の結果を Fig.3 に示す。クラスタをそれぞれ見ると、関連した単語が集まっていることが分かる。例えば、造語や人物名がまとまっていたり、汎用的な単語がまとまっていたりしている。次に、シーン分割の結果として得られたシーンの単語クラスタリング結果の一例を Fig.4 に示す。こちらはクラスタごとに見てみると発言ごとの単語のまとまりになっており、単語の類似性が発言内の方が高いことがわかる。また、どちらの発言においても共通の単語についてはまた別のクラスタとして分けられていた。

## 5.2 身体動作の同調の分析結果

4.3 で述べた同調検出の結果の一例として 1.0 ~ 5.0 [Hz] 帯を 1.0 [Hz] ごとに示したものを Fig.5 にまとめた。今回使用した分析データは伝達の内容から共感の内容の順番で説明が行われたが、どの周波数帯においても後半部分に同調現象が多く見られた。これは井上の報告<sup>5)</sup>でも見られた現象で、実験データから再現ができていたことが示せた。

## 5.3 発話内容と身体動作の同調の関係

発話内容の分析ではシーンの分割ができ、身体動作の同調においても井上の分析結果<sup>5)</sup>の再現を行うことができた。そこで、分割されたシーンの時間区間ごとに身体動作の同調度を総計することで、どのシーンにおいて多く同調が起きているかを調べた。なお、シーンごとに時間の長さが違うため、シーンごとの身体動作の同調度はそのシーンに含まれる複数発言の始まりから終わりまでの発言時間を割ることで標準化した。結果の一例として 1.0 ~ 5.0 [Hz] 帯を 1.0 [Hz] ごとに示したものを Fig.6 にまとめた。全体を通して見ると 1 ~ 2, 9 ~ 10 シーン目の間、特に 10 シーン目に身体動作の同調が見られた。ここで、シーン分割における最初の分割箇所がどこかを確認した。すると、9 ~ 10 シーン目の間が最も話題の変化が大きい箇所であることが分かった。つまり、話題が大きく変わった時や話を始めた時に身体動作の同調が多く発生していたことになる。

## 6. 考察

本研究では、坂原の手法<sup>7)</sup>を用いた全発言におけるシーン分割と井上の手法<sup>5)</sup>を用いた身体同調の分析を行い、分割されたシーンごとに身体動作の同調度を調べた。その結果として、シーン 1 ~ 2, 9 ~ 10 の間で身体同調が多く発生していることがわかった。この結果から、発言内容と身体動作の同調との関係性の分析手法を確立することができたとと言える。

Table.1 Speeches of each scenes.

Scene	No.	Contents
1	1	これから、先人が先人が考え、学問的に考えた、人生の意義について説明します。
	2	えと、人生の意義は、人生において目的や意味はあるのか、あるとすれば、それはいかなるものか、考えたってことです。
2	3	で、哲学の分野ではこの問いに対して多くの議論がなされてきました。
	4	例えば論理実証主義者は、人生の意味とは何かと、でえーと、そのことについて問うことに意味があるのか、と問いかけたことがあります。
	5	で、これはもし客観的な価値が存在しないとすれば、人生は無意味なのだろうか、という考えです。
3	6	で、これに対してルードヴィッヒヴィットゲンシュタインという、と、論理主義、論理実証主義者たちは言語によって表現されるならば、その問いは無意味であると定義しました。
	7	また、功利主義の創始者であるジェレミーベンサムは、快と不快という二つの支配、支配者のもとにあることが、人間の自然であると主張し、そして、道徳的洞察から功利性の支配という説を展開し、善は何であれ、最大多数の最大幸福である、としました。
	8	その上で彼は生きる意味を最大幸福の原理として定義しました。
4	9	一方で、ニヒリズムという考えは人生には意味はないと定義しました。
	10	一方で、ニヒリズムという考えは、
6	11	あ、違う間違えた。
	12	ニヒリズムというのは、最も高い価値があるものを、無価値にしてしまう、過程と言えます。
7	13	フリードニヒニーチェはニヒリズムの特徴というのは世界、人間の意味、存在、目的、可知的真理、本質的価値などを空っぽにしてしまうことだ、と述べました。
	14	また、マルティンハンデガーはニヒリズムというのは存在が忘れ去られてしまい、存在を価値へと変容させてしまう動きであり、価値へと交換するために、存在を減らしてしまうことである、としました。
	15	このように、人間人生の意義についての問いは多くの哲学者、議論と多くの思想を生み出してきました。
8	16	ここで、実際の、現実社会においては、この人生の意義という問いは、経済的に豊かな国ほど切実な問題になる、なります。
	17	経済的、物質的に豊かな国はひどい空虚感や心の虚しさに苛まれている人の数が増える傾向にあります。
9	18	この理由として、人間は衣・食・住といった基本的欲求のすべてを満たしてようやく自己実現の欲求といった高次欲求に駆られ始める、と心理学者は主張します。
	19	豊かな社会は基本的欲求を満たしやすい社会なので、高次よっ、高次の欲求が発現しやすく、それが満たされない苦しみにさいなまれやすいって事です。
	20	また人生においてこのような命題が人の心をとらえる時期が三つあります。
10	21	それが思春期、中年期、老年期の三つです。
	22	で、思春期を得たものの多くはその段階なりに解答を持ちます。
11	23	中年期あるいは老年期にも、このような心、このような問いが心とらえることがあります。
	24	んで、これは若い頃に財産、地位、家族などを手に入れば幸福になるに違いないと思っていって、ひたすら頑張ってきたのに、いざそれらを手に入れてみると、全く幸福という実感がなく、じんし、自分の人生に大切な何か欠けている、という気がして仕方なくなり、人生のむなしさを痛感することです。
	25	この段階で、改めて、残された人生は意味のあるものにするために今後どう生きていけばいいのだろう、と、言う問いに真正面から向き合うことになります。
12	26	この問いはそもそも、自身の価値観の決定、或いは態度の決定に関する問いであるので、学問や化学はこの問いに対する解答を与えてくれないと社会学者は主張しています。
	27	以上で、説明終わります。

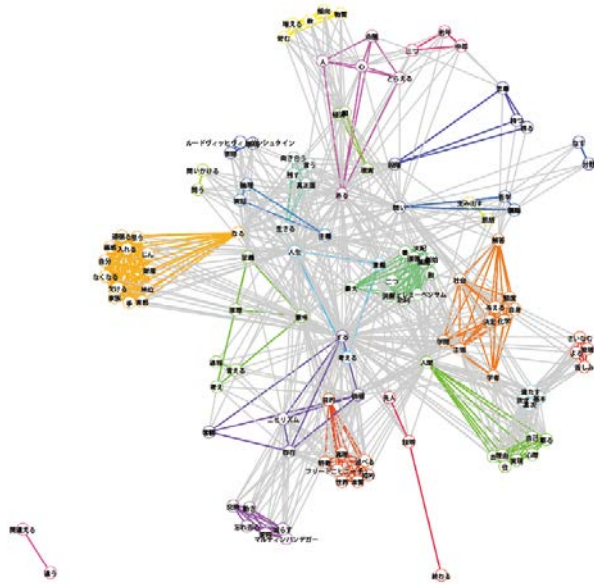


Fig.3 Cluster of all speeches.

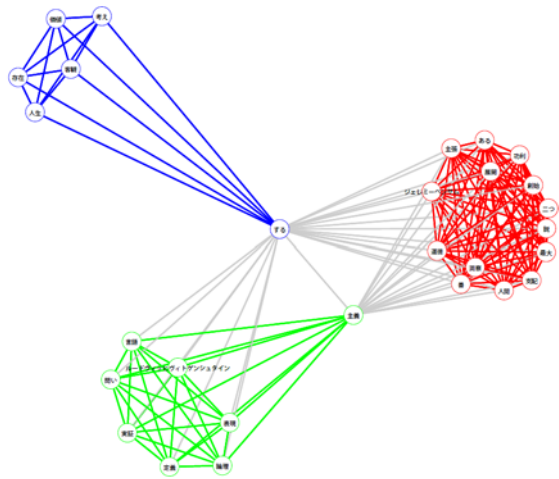
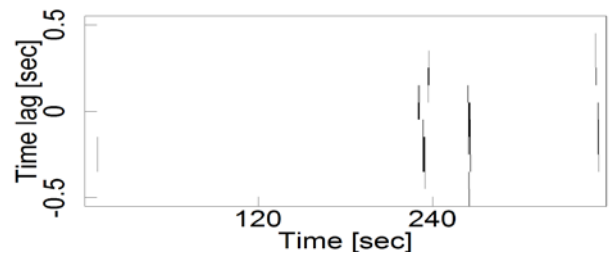


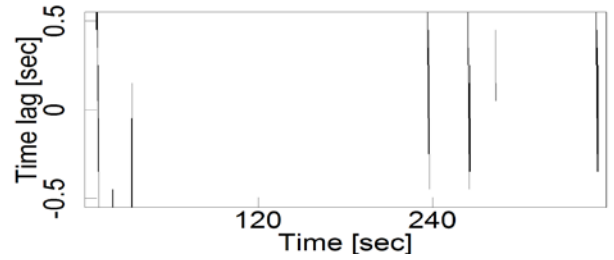
Fig.4 Cluster of 5 ~ 7 speeches.

また、話題が大きく変わった時や話を始めた時に身体動作の同調度が多く見られたことから、発言内容と身体動作の同調との間に関係性があることが示唆された。

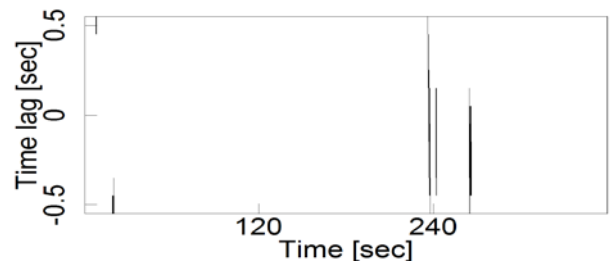
そこで今後の課題として、発言内容が身体同調とどのように相関しているかを調べるために、発言内容の構造やシーンごとの特徴などを分析することがある。例えば、シーン分割の度に行う単語クラスタリングの結果を階層的に見ることで、発言内容の話題の推移や話題間の関係性を見ていく<sup>[12]</sup>。また、Word2Vec といった技術<sup>[13-14]</sup>を用いて各シーン内の単語やフレーズなどに特徴付けを行うことで、シーンごとの違いを調べる。



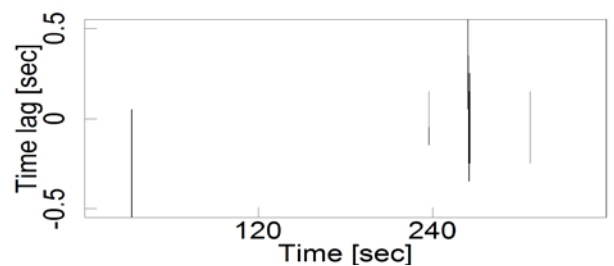
(a) 1.0 [Hz]



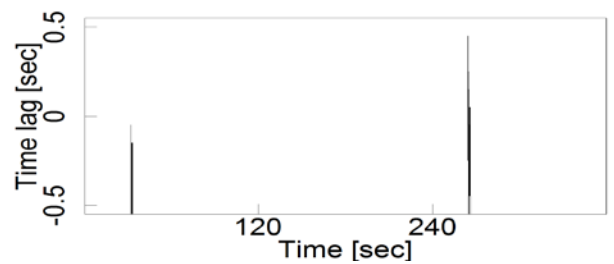
(b) 2.0 [Hz]



(c) 3.0 [Hz]



(d) 4.0 [Hz]



(e) 5.0 [Hz]

Fig.5 Degree of synchronization of each frequency.

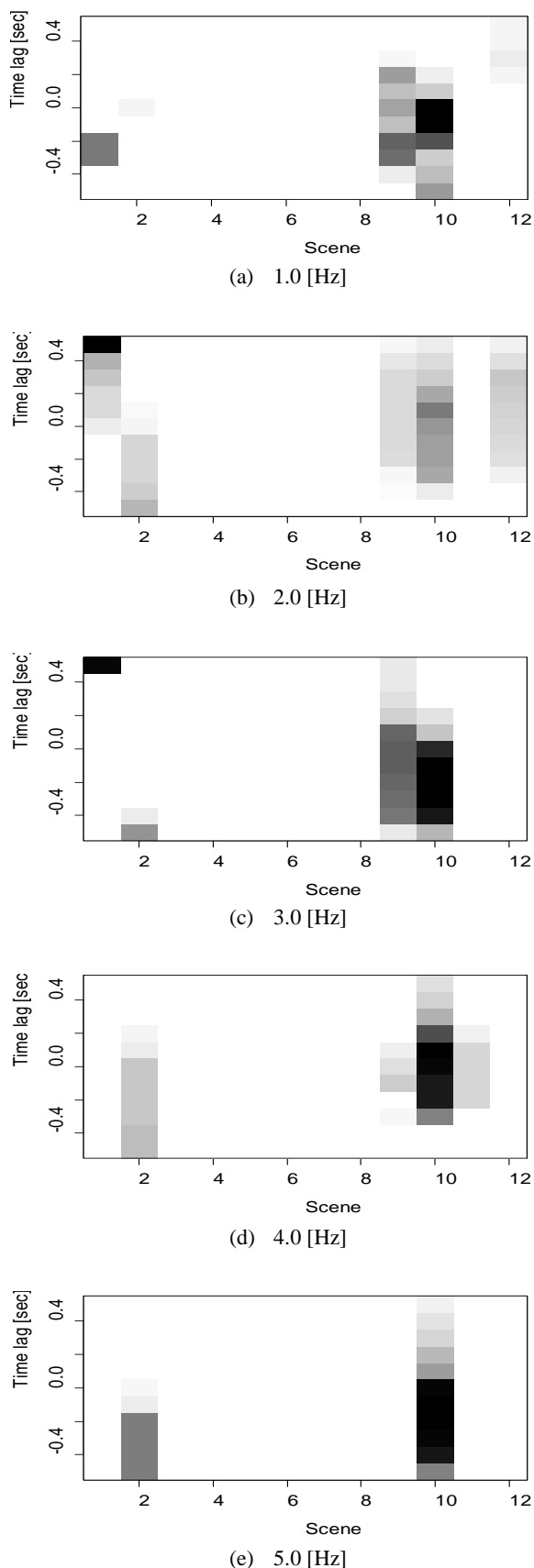


Fig.6 Degree of synchronization of each scene in each frequency.

## 7. 結論

本研究では、人間の対話コミュニケーションにおける言語・非言語チャンネルの関係性を分析することを目的として研究を行った。先行研究の分析手法を<sup>[5][7]</sup>をもとに言語・非言語チャンネルの関係性を分析する手法を確立し、分析を行った。その結果から、両者に関係性があることが示唆された。今後、各シーンにおける特徴を抽出し、身体動作の同調との関係性を調べたいと考えている。

## 参考文献

- [1] 梶本：人間関係からみた課題解決の会話の連鎖構造；世界の日本語教育，日本語教育論，No.10, pp.221-239 (2000).
- [2] 松本，他：感情生起事象文型パターンに基づいた会話文からの感情推定手法；自然言語処理，**Vol.14**, No.3, pp.239-271 (2007).
- [3] 塩崎，他：設計における談話の分析と構造化；電子情報通信学会技術研究報告，KBSE，知能ソフトウェア工学，**Vol.97**, No.632, pp.41-48 (1998).
- [4] Ono, E., *et al.*: Relation between synchronization of head movements and degree of understanding on interpersonal communication; Proc. of the 2012 IEEE/SICE International Symposium on System Integration (SII), pp.16-18 (2012).
- [5] 井上：合意形成過程における身体動作の同調現象の分析；修士論文，東京工業大学大学院総合理工学研究科知能システム科学専攻 (2014).
- [6] Wikipedia: 人生の意義；<https://ja.wikipedia.org/wiki/%E4%BA%BA%E7%94%9F%E3%81%AE%E6%84%8F%E7%BE%A9>
- [7] 坂原：TV番組の字幕情報を利用したシーン分割とアノテーションの生成；修士論文，東京工業大学大学院総合理工学研究科知能システム科学専攻 (2014).
- [8] Kudo, T., *et al.*: Applying Conditional Random Fields to Japanese Morphological Analysis; Proc. of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.230-237 (2004).
- [9] MeCab: Yet Another Part-of-Speech and Morphological Analyzer; <http://taku910.github.io/mecab/>
- [10] Frey, J. B. and Dueck, D.: Clustering by passing messages between data points; Science, **Vol.315**, No.5814, pp.972-976 (2007).
- [11] Inoue, Y., *et al.*: Head motion synchronization in the process of consensus building; Proc. of the 2013 IEEE/SICE International Symposium on System Integration (SII), pp.70-75 (2013).
- [12] 斎藤，片上，新田：データ結晶化を用いた対話ログの時系列解析；情報処理学会研究報告，**Vol.2010-CE-103**, No.2, pp.1-8 (2010).
- [13] Mikolov, T., *et al.*: Efficient estimation of word representations in vector space; Proc. of the International Conference on Learning Representations 2013 (ICLR), arXiv preprint arXiv:1301.3781 (2013).
- [14] Mikolov, T., *et al.*: Distributed representations of words and phrases and their compositionality; Proc. of the 2013 Conference on Neural Information Processing Systems (NIPS), pp.3111-3119 (2013).