

## 対面および非対面における対話の違い ～頭の振りの役割分析～

岩野裕利 杉田洋介 松永美穂 白井克彦

早稲田大学理工学部

〒169 新宿区大久保 3-4-1

E-mail: iwano@shirai.info.waseda.ac.jp

人間の対話において、身振り、表情、頭の振りと言った視覚情報は、より円滑な対話を行なう上で重要な情報である。より自然なヒューマンインターフェースを構築する上でも、音声だけではなく、これらの視覚情報をどのように利用していくかが大きな課題となっている。本研究では、人間同士の対面および非対面の対話を比較することにより、視覚情報の中でも頭の振りに注目し、対話における役割に関して分析を行なった。分析の結果、頭の振りには、情報伝達手段として利用される頭の振り、と、そうでない2種類の頭の振りがあることを確認した。また頭の振りが対話における発話のタイミングと関係していることが示唆された。

## Difference in Face-to-face and Telephone Dialogues: Analysis of the Role of Head Movement

Yuri Iwano, Yosuke Sugita, Miho Matsunaga and Katsuhiko Shirai

School of Science and Engineering, Waseda University

3-4-1 Okubo, Shinjuku, Tokyo 169, Japan

E-mail: iwano@shirai.info.waseda.ac.jp

In practical conversations, visual information such as gesture, facial expression and head movement clearly makes the progress of conversation much smoother and more natural. Therefore, in the more natural human interface that can use multiple modalities, visual information becomes as important as voice information. In this research, we analyzed conversations between face-to-face and conversation through telephone line. It seems that there are two types of head movements, depending on whether it is intended to give an information to his partner or not. Also head movements seem to have a correlation between the timing of utterances.

## 1 はじめに

対話は主に音声による双方向の情報伝達である。実際、電話における会話を考えてみれば、対面だけでなく音声のみで円滑な対話が可能である。しかし実際の会話においては、聴覚情報だけではなく、頭の振り、身振り、表情などといった視覚情報が、我々の会話をより円滑に進めていると考えられる。また、対話の分析を行なう上でも、音声情報のみでは扱う事ができない対話における聞き手側の情報についても触れることもでき、より詳細な対話分析が可能となる事が期待される。複数のモダリティーが利用できる、より自然なヒューマンインターフェースの構築において、音声情報だけではなく視覚情報の利用が重要となってくる。

視覚情報が我々の対話をより円滑なものとする要素の1つであることは明らかである。本研究では、視覚情報の中でも頭の振りに注目する。非言語的行動には標準的な規則がなく、文化的な違い、個人差も大きい。我々人間の動作の中で、頭の動きは動作がはっきりしており分類することが容易である。また英語圏の会話と比較して日本語における会話では頭の振りは比較的多く出現することが知られている[1]。よって、個人差があっても、頭の動きはある程度まとまった量のデータを分析することにより、統計的な傾向が得られる視覚情報の一つである。

多くの対話分析に関する研究は、音声情報を主としたものであり、対話における詳細な視覚情報の分析はあまり行なわれていない。本研究では、対面および非対面の2つの条件下での対話を分析した。対面および非対面の対話を比較することにより頭の振りが対話においてどのような役割を果たしているのかを中心に分析し、視覚情報の対話における重要性について述べる。

## 2 対話収録

### 2.1 対話実験

本研究では、2つの異なった条件下における対話を分析した。視覚的に情報の伝達の可能な対面および、不可能である非対面の対話実験である。対面実験の様子を図1に示す。対話における視覚情報の役割を分析するためには、情報の伝達が音声情報のみで成り立っている対話を分析する必要がある。実験

環境の被験者に与える不自然さを極力抑えることが重要であるので、我々の生活に密着した電話を利用することによって、情報の伝達が音声のみで行なわれる環境を模擬した。実際には、被験者は同じ部屋にいますが、お互いに相手を見ることはできない(図2)。

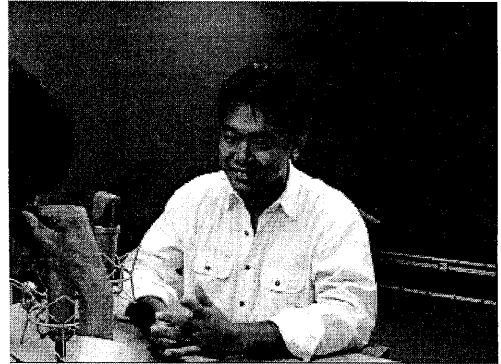


図 1: 対面の実験の様子

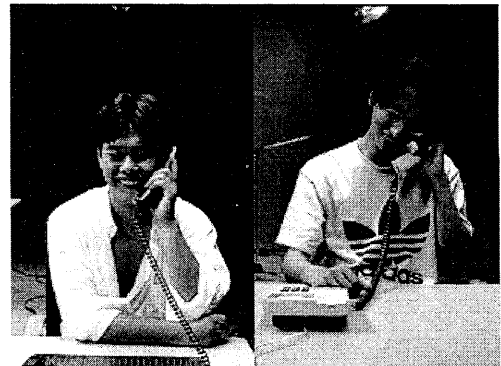


図 2: 非対面(電話)の実験の様子

お互いを知っている10名の学生、合計5組にお願いし、対面および非対面の対話を収録した。収録は、早稲田大学理工学部のマルチメディアスタジオを利用した。被験者に与えたタスクは、4コマ漫画のシナリオを考えるというものである。4コマ漫画の内容を詳細に決めるだけではなく、タイトルとキャッチコピーも決めて貰う。

対話の様子は、2台のカメラによって動作を、また2本のマイクによって、DATに様々な音声処理に適した高音質(16bit 48KHz)な状態で録音した。

DATの各チャンネルに、それぞれの話者の発声が収録されている。ただし、2本のマイクはそれぞれの話者の方を向いているが、相手のマイクに発声が入ることもあり、発話を完全に分離できていないわけではない。

同じ組の被験者が対面および非対面の対話実験を異なる日に行なった。また、実験の順序による影響を避けるために、対面と非対面のどちらを先に行なうかは、全被験者の間でバランスをとった。対話における制限は何も付けず、被験者は好きなことを好きなだけ話せるものとした。4コマ漫画のシナリオ、タイトル、キャッチコピーが決まった段階で実験は終了する。対話の長さは、15分から41分にわたり、対面の総対話時間は約117分、非対面は約93分であった。

## 2.2 対話データ

全ての対話の音声は、ワークステーションにデジタル転送を行なった。それぞれの対話は、1つのファイルとして転送されており、対話の時間情報が扱えるようになってきている。対話内容の書き起こし専用ツールを用いて、発話者、その発話の開始時刻、終了時刻そして発話内容がファイルとして保存されている。同一の話者の発話でも、発声と発声の間にポーズがある場合は違う発話として扱った。発話には、ラベル付けツールを用いて対話の状態を示すラベルと発話の機能的な内容を示すラベルが付加されている。頭の動きは、ビデオより人間の目によって抽出を行なった。本研究で使用した対話データは、頭の振りの分析のみだけでなく様々な対話分析が可能となっている。図3にデータの構成を示す。対話における聴覚と視覚情報を統合的に分析する一端として本研究は行なわれている。

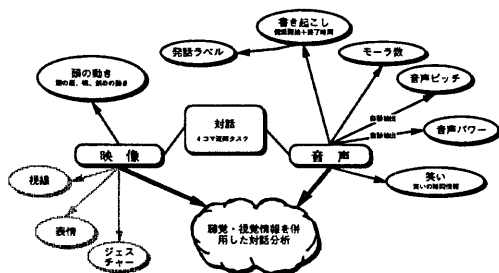


図3: 対話データの構成

## 3 頭の振りの分析

我々の会話において、うなずきのような頭の振りは、非対面、つまり相手が見えないような環境においても確実に起こる。よって、頭の振りは必ずしも相手に情報を伝達するためのものではないと言える。ここで、頭の振りには2種類のタイプがあると推測できる。1つは、述べたように、無意識あるいは自分の理解を確認するためのもの、そしてもう1つは、相手に情報を伝えるための動きである。

対面および非対面の対話を分析することによりこれらの2種類の頭の振りを分類し、違いについて考察する。本研究における頭の振りとは、頭の縦、横、斜め（傾げる）の振りで、体全体の動きや、視線を変えることによって伴う頭の動きは無視した。

## 4 結果

対面および非対面における頭の振りに関して、それぞれ5対話づつ分析を行なった。表1に頭の振りの出現頻度および、その振りが発声を伴ったものかどうかという観点での分類を行なった。予想されるように、対面と非対面の総対話時間の違いを考慮しても、頭の振りはお互いに相手を見ることのできる対面の対話においてより多く現れた。また、非対面の5.6%に対して、対面における頭の振りの約18%には発声が伴わなかった。

表1: 頭の振りの出現頻度

分類	出現頻度		
	発声	対面	電話
縦	有	1451	673
縦	無	314	42
横	有	21	4
横	無	3	0
斜	有	24	5
斜	無	12	0
合計		1825	724

表2に、発話を伴わない頭の振りが、対話相手の発話中に現れたものなのか、相手の発話終了後に現れたのか（図4）で分類し、その出現頻度を示す。

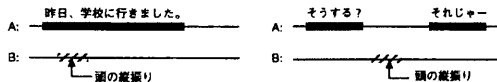


図 4: 発声を伴わない頭の振りの分類の説明

表 2: 発声を伴わない頭の振りの分類

頭の振り	相手が発話中		発話終了後	
	対面	電話	対面	電話
縦	85	25	229	17
横	1	0	2	0
斜	3	0	9	0
合計	89	25	240	17

表2より対面の場合、発話終了後に発声を伴うことなく、情報伝達手段として頭の振りがより多く利用されていることが確認できる。頭の振りが相手に情報を伝えることができることを考慮すると、この結果は音声情報を伴わなくても頭の振りが対話に影響を与えることが可能であるという、視覚情報の重要性を示唆するものである。

ここで、頭の振りのより詳細な分類を定義する。頭の振りはその動作の意味に応じて分類することができる。表3に示す分類を定義し、頭の振りをその意味に応じて分類しその出現頻度を表4に示す。

表 3: 頭の振りの意味ラベル

ラベル	説明
発話要求	相手に発話を要求するような発話の最後に伴った頭の振り
同意	明白な肯定、同意を示すもの
否定	否定を表すための頭の振り
あいづち	あいづちに伴う頭の振り
復唱	相手の発話の復唱に現れる頭の振り
繋語	発話の間を繋ぐ発話に伴う頭の振り
強調	発声を強調するための頭の縦振り
懷疑	懷疑的な気持ちを表す頭の振り

ほとんどの頭の横振りは相手に否定的な意思を伝えるためのものであり、また斜めについても懷疑的な意思を伝えるためのものである。横および斜めの

表 4: 頭の振りの意味ラベルの出現頻度

頭の振りの意味	対面			電話		
	有*	無	合計	有	無	合計
発話要求	322	0	322	121	0	121
同意	248	43	291	143	8	151
否定	19	3	22	5	0	5
あいづち	389	275	664	210	34	244
復唱	32	0	32	14	0	14
繋語	34	0	34	29	0	29
強調	429	0	429	156	0	156
懷疑	23	8	31	4	0	4
合計	1496	329	1825	682	42	724

\*発声を伴っているかないないか

頭の振りの意味は比較的是っきりしており、また出現頻度が低いので本研究においては、頭の縦振りにのみ注目する。

頭の振りの意味ラベルの中で、「あいづち」、「同意」、「否定」、「懷疑」に関しては、音声情報を利用することなく頭の振りで同等の情報を伝達することができるが、その他のラベルに関しては、必ず発声に伴う性質のものである。表4より、発声を伴わない頭の振りは、特に「あいづち」の際によく現れる。対面の対話においては、41.4%(275/664)のあいづちを意味する頭の振りに発声が伴っていない。「同意」に関しても、14.8%(43/291)の頭の振りに発声が伴っていない。相手が必ずしも頭の振りを見ているとは限らないという問題もあるが、少なくとも発声と伴わない頭の縦振りは、視覚情報のみが情報伝達手段であることは確かである。

#### 4.1 頭の振りの出現割合

ここで、発話に頭の振りを伴う割合についての分析を行なう。表6に、対話における発話の機能ラベルとその出現頻度を示す。機能ラベルとは、対話中の全ての発話に関して付加されており、対話における発話の機能を示すものである。ここでの発話の機能ラベルは、ある程度まとまった発話に関して付加されているものである。一方、頭の振りの意味ラベルに関しては、頭の振りの現れた局所的な発話に関して付加されている。よって、必ずしも発話の機能

ラベルと頭の振りの意味ラベルが1対1で対応しているとは限らない。当然のことながら、発話の機能ラベルは書き起こし（発話内容）に対して付加されているものであるため、発声の伴わない頭の振りの意味ラベルとの対応はない。

表 5: 発話の機能ラベルの定義

ラベル	説明
発話要求	聞き返し、疑問文、呼掛けのように相手に何らかの応答を要求する発話
同意	明白な同意の意思を示すもの
否定	明白な否定の意志を示すもの
返答	同意、否定に分類できない返答
あいづち	相手の発話に調子を合わせる発話
復唱	相手の発話の復唱
繋語	発話の間を繋ぐためのもの
独語	独り語
情報提示	上記の分類に該当しない発話で情報を提示している発話

表 6: 発話の機能ラベルの出現頻度

機能ラベル	対面	電話
発話要求	855	714
同意	408	392
否定	47	37
返答	114	75
あいづち	599	568
復唱	145	93
繋語	337	458
独言	30	27
情報提示	1569	1107
合計	4104	3471

表 6より、「繋語」に関しては、非対面の対話の方が多く出現している。「繋語」は、発声を伴わないと成り立たないものであるため、これは相手が見えない事による対話上の特性の違いと言えるであろう。また、対面の総対話時間が非対面のと比較して1.25倍長いので、単純には比較できないが、「あいづち」、「同意」に関しても対話時間の長さの違いを考慮すると、非対面の対話の方により多く現れる傾向にある。「繋語」の場合と異なり、「あいづち」、「同意」に関しては発声を伴わなくても頭の振りで

同等の情報を伝達することが可能である。表 4からも分かるように、情報の伝達を音声をあえて使わずに、頭の振りによって代用していることが、表 6における対面における「あいづち」や「同意」の発話の機能ラベルの出現頻度の減少に貢献していることが確認できる。一方、表 7より、非対面の対話においてはいずれの場合においても、頭に振りを伴う発話が出現する割合は対面と比較して低い。

図 5から図 9まで、あいづち、同意、発話要求、復唱、繋語における発話を伴う頭の振りの出現確率を示すものである。強調に関しては、対応する発話の機能ラベルがないので、分析からは除外する。

図における各点は、各話者の頭の振りの出現確率を意味する。ダイヤモンドは、平均値と95%信頼幅を示し、矩形は分位値を示しており、上から順番に10%、25%、50%、75%、90%のパーセンタイル値を示している。

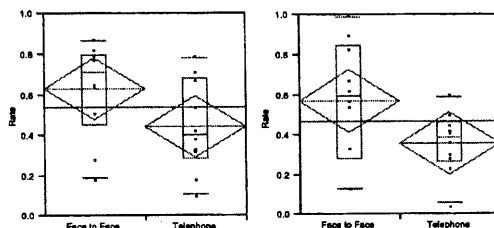


図 5: あいづち

図 6: 同意

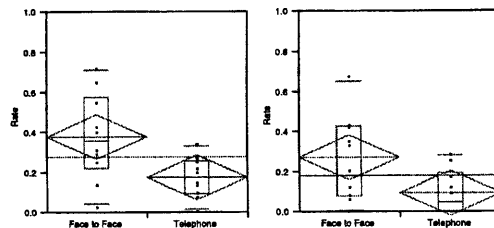


図 7: 発話要求

図 8: 復唱

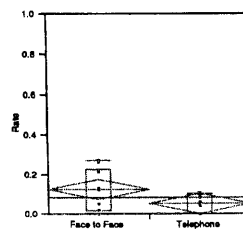


図 9: 繋語

表 7: 頭の振りの現れる確率 (被験者 10 名の平均)

ラベル	対面		電話	
	平均	標準偏差	平均	標準偏差
あいづち	63%	0.24	44%	0.23
繋 語	12%	0.10	5%	0.04
同 意	57%	0.30	36%	0.16
発話要求	38%	0.21	18%	0.10
復 唱	27%	0.21	9%	0.11

表 7は、被験者 10 名の頭の振りの出現確率の平均と標準偏差を示すものである。これらの図より、発声に頭の縦振りが伴う割合は相手が見えない電話実験において低くなることが確認できる。更に、対面における頭に縦振りの出現確率の話者によるばらつきが非対面と比較して大きくなっている。相手が見えることによって増加する頭に振りに関して、より個人差が現れるものと考えられる。

#### 4.2 頭の振りの対話の遷移に与える影響

次に頭の振りが対話の遷移に影響を及ぼすかについて分析を行なう。相手に何らかの応答を求める「発話要求」を意味する発話に注目し、その発話に頭の振りが伴っているかないかで、返答を得る割合が変わるかを調べる。表 8に対面および非対面における発話要求に対して何らかの返答(同意、否定、答えなど)あるいは全く関係ない発話が続く割合を示す。

表 8: 頭の振りの対話の遷移に与える影響

頭の振り	対面 (%)		電話 (%)	
	返答	その他	返答	その他
有り	61.3	38.7	62.0	38.0
無し	50.7	49.3	50.4	49.6

表 8より、一見すると、頭の振りを伴う発話要求は伴わないものと比較してより多く返答を得られるように思われる。しかし、対面と非対面における特徴がほとんど同じことより、発話要求の後の対話の遷移に関しては、視覚情報の影響はほとんどないと言える。

しかし、発話要求の後の発話に関して分析を行なっただけであり、より詳細な対話の遷移の分析

を Ergodic HMM[5] などを利用して行なう必要がある。

#### 4.3 オーバーラップの分析

ここでは、頭の縦振りが発声のタイミングに利用されているかに関して分析を行なう。表 9に頭の縦振りを伴った発話の次発話とその発話にオーバーラップした回数を示す。

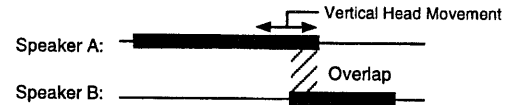


図 10: オーバーラップの状態

表 9: 頭の縦振りを伴った発話に対して次発話がオーバーラップする割合

	対面	電話
出現回数	280	88
(頭の縦振りの全出現回数)	(1451)	(673)
割合	19.3%	13.1%

表 9より、頭の縦振りが伴った発話に対して次発話がオーバーラップする割合は、対面において 19.3%(280/1451)、非対面において 13.1%(88/673)であった。相手の見える対面において非対面と比較してより多く現れている。対面および非対面における出現割合の大きな違いはないが、対面において、頭の縦振りとオーバーラップの間に何らかの関係があると言えるだろう。我々は、相手の頭の縦振りを利用してオーバーラップを行なうタイミングを得ていると考えられる。

一方視点を変えて、表 10に本研究に用いた対話の時間情報を示す。表の中の数字は、それぞれの項目の対話中における時間の割合を示す。発話割合は、それぞれの話者の発話時間の割合で、同時発話は双方の話者が同時に発声している時間の割合、沈黙は双方の話者ともに発声していない時間の割合を示す。つまり発声 A + 発声 B + 沈黙 - 同時発話 = 100 となる。この表より、同一の組の対面と非対面を比較すると、沈黙時間の割合は似通っており、

対面、非対面による発話の割合には違いがほとんど現れない。一方、オーバーラップの割合は相手が見える対面において増加する。対話をしている2名の話者の発話時間の割合が均等な場合この傾向は強く出ている。話者のいずれかが、より多く話すような場合（EF組、IJ組）、オーバーラップの割合は話者性的影響をより大きくうけていることが考えられる。対面の対話においてオーバーラップして発話している時間が長くなるという傾向を確認した。

表 10: 対面と非対面の比較(対話時間に対する割合)

話者	対面 (%)			非対面 (%)		
	発話割合	同時発話	沈黙	発話割合	同時発話	沈黙
A	28.6	2.9	47.9	21.2	1.5	47.6
B	26.5	2.9	47.9	32.7	1.5	47.6
C	27.7	2.3	46.0	29.3	1.2	45.4
D	28.6	2.3	46.0	26.5	1.2	45.4
E	36.6	4.8	43.2	37.9	4.6	43.6
F	25.0	4.8	43.2	23.0	4.6	43.6
G	30.7	3.9	40.3	31.2	2.7	43.2
H	32.9	3.9	40.3	28.4	2.7	43.2
I	37.4	3.8	39.3	41.5	4.0	31.3
J	27.2	3.8	39.3	31.3	4.0	31.3
total	-	3.7	43.6	-	2.6	42.3

#### 4.4 発話のタイミング

対話のリズムを形成する要素として、発話の内容、スピード、ポーズなどが上げられる。これらの要因は、互いに関係し合っており切っても切り離せないものであるが、ここでは、ポーズの対面と非対面の対話における違いについて述べる。

人間はポーズの違いを意識せず対話相手と自然にタイミングを合わせて対話を進めている。このことから人間は対話のリズムに合わせて、発話のタイミングを変えている可能性があると考えられる。そこで対話において話者が互いに発話のタイミングを合わせていく傾向があるかどうかを調べるために、話者が交替する際の話者毎の各発話の機能ラベル(表5)の前のポーズ長の平均を求める。そして、この平均長が全話者の平均より長いか短いかで分類を行

なう。この結果を用いて、各対話の組の中で、話者2名が同様のポーズ長の傾向になるかを、対面と非対面の両方で分析した。

表 11、表 12にそれぞれ対面、非対面の対話における結果を示す。表の中の「L」は、話者2名のポーズの傾向が平均より長いということで一致、「S」は短い傾向が一致した場合、「×」は一致しなかった場合、また「-」は該当するラベルのデータがなかったため分析出来なかったことを示す。

表 11: 対面における平均ポーズ長の傾向の比較

ラベル	対話の組					同一の割合 (%)
	1	2	3	4	5	
発話要求	L	S	L	×	×	60 (3/5)
同意	L	S	L	S	×	80 (4/5)
否定	-	S	-	-	S	100 (2/2)
返答	L	×	×	S	S	60 (3/5)
あいづち	L	S	L	S	S	100 (5/5)
復唱	S	×	L	×	×	40 (2/5)
繋語	L	S	×	S	S	80 (4/5)
独言	-	-	×	-	-	0 (0/1)
提示	L	S	L	S	S	100 (5/5)
合計						73.7 (28/38)

表 12: 非対面における平均ポーズ長の傾向の比較

ラベル	対話の組					同一の割合 (%)
	1	2	3	4	5	
発話要求	L	×	×	×	S	20 (2/5)
同意	L	×	×	S	S	60 (3/5)
否定	L	S	-	×	S	75 (3/4)
返答	L	×	×	S	S	60 (3/5)
あいづち	L	S	×	×	S	60 (3/5)
復唱	S	×	S	×	×	40 (2/5)
繋語	S	S	L	L	S	100 (5/5)
独言	-	-	-	×	-	0 (0/1)
提示	×	×	L	L	S	60 (3/5)
合計						60 (24/40)

このポーズ長の分析には、前発話の内容による影響や、ポーズ長の平均より長いか短いかという比較

方法の曖昧性などと言った問題点はあるが、傾向を捉えるという観点で考察を行なう。

表 11、表 12 より、ポーズ長の傾向が同じになる割合が対面の対話においては 73.7%、非対面の対話においては 60% となった。対面と非対面の一番大きな違いは、視覚情報があるかないかであるが、ポーズ長の傾向が同じになる割合が対面の方が多いということは、発話のタイミングの決定要素の 1 つとして視覚情報が利用されていることが示唆される。

対面および非対面で、両方とも比較的ポーズ長の傾向が一致する発話の機能ラベルは「繋語」の場合である。「繋語」の定義は発話の間を繋ぐためのものであるので、発話中、相手に対話の主導権が移動するのを防ぐためなどに、その対話において発声に対話の流れに沿ったタイミングで現れる傾向にある。このことを考慮すると、ある対話のリズムと「繋語」の発声タイミングが一致する傾向になることはうなずける傾向と言える。

また、ポーズ長の傾向が非対面と比較して対面の方で一致する傾向にあるものの中で、「発話要求」、「あいづち」、「同意」に関しては、表 4 の頭の振りの意味ラベルの出現頻度および、表 7 の頭の振りの出現割合からも、これらの発話に頭の振りが伴う割合の多い発話であることがわかる。つまり、視覚情報のある対面で、ポーズ長の傾向がより一致するという傾向と、一致する傾向にある発話には頭の振りが伴う割合が多いということが結び付く。

視覚情報と発話のタイミングには相関があることが示唆されたが、さらに詳細な分析が必要である。例えば、1 名の話者を固定で複数の相手との対話などといった分析を行なうことにより、より詳細な対話のリズムに関する分析ができることが期待される。

## 5 おわりに

対面および非対面の対話における頭の振りの分析を行ないその違いに関して述べた。

相手に情報を伝えるための頭の振りかどうかという観点で 2 種類の頭の振りのタイプを仮定した。頭を振っている人の視点に立つと、頭の振りの一部は相手に何らかの情報を伝えるためのものである。例えば、もし同意の意思を音声と頭の縦振りの両方で表現した場合、相手によって知覚された情報は主に音声情報によるものである。一方、発声を伴わな

い頭の振りが存在することより、得られた情報が主に視覚情報によるといった逆のケースも存在する。

対話相手の頭の振りを見ている側の人間の視点に立つと、相手の肯定や否定などといった意味を持つ頭の振りを知覚することにより、情報を得るだけでなく、頭の振りは相手の発話に対してオーバーラップをするタイミングを得るのに利用されているようである。また、発話のポーズの分析より視覚情報が発話のタイミングと関係があることが示唆された。

頭の振りを、信号としてだけではなく双方向の情報伝達手段として捉えることにより、対話における視覚情報の影響を見ることが出来る。より有効なマンマシンインターフェースを構築する上でも、音声情報と共に、視覚情報の利用が重要になってくる。また、対話における発話の役割、動的な構造を分析する上でも頭の振りは有効な要素となると考えられる。

## 参考文献

- [1] 泉子 K. メイナード: “会話分析”, くろしお出版, 1993
- [2] 坂本憲治、綿貫啓子、外川文雄: “マルチモーダル対話解析”, 人工知能学会研究会資料, SIG-FAI-9401-6(6/6), 1994.
- [3] 綿貫啓子、坂本憲治、外川文雄: “マルチモーダル対話データベースにもとづく対話解析”, 情報処理学会音声言語情報処理研究会, 5-3, (1995.2.3).
- [4] 綿貫啓子、外川文雄: “対話における感情の变化の解析”, 情報処理学会音声言語情報処理研究会, 7-13, (1995.7.21).
- [5] Katsuhiko Shirai: “Modeling of Spoken Dialogue with and without Visual Information”, Proc. of ICSLP96, vol1, pp188-191, 1996.
- [6] Yuri Iwano, Shioya Kageyama, Emi Morikawa, Shu Nakazato, Katsuhiko Shirai: “Analysis of head movements and its role in spoken dialogue”, Proc. of ICSLP96, vol4, pp.2167-2170, 1996.