The University of Osaka
Institutional Knowledge Archive

Title	二者間会話場面における非言語行動の抽出法 : 動画・音声解析ソフトとイベントレコーダーについての比較検討
Author(s)	藤原, 健; 大坊, 郁夫
Citation	対人社会心理学研究. 2012, 12, p. 165-172
Version Type	VoR
URL	https://doi.org/10.18910/6664
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

二者間会話場面における非言語行動の抽出法 1)

—動画・音声解析ソフトとイベントレコーダーについての比較検討—

藤原　健（大阪大学大学院人間科学研究科・日本学術振興会）
大坊郁夫（大阪大学大学院人間科学研究科）

本研究の目的は、二者間会話場面における非言語行動の抽出法について新たな抽出法を採用し、その妥当性および有用性について検討することである。具体的には、これまでのイベントレコーダーを用いた観察による抽出法と近年注目されつつある動画・音声解析ソフトを用いた抽出法により、ジェスチャー、アダプター、発話の 3 つの非言語行動を抽出した。そのうえで、これら 2 つの抽出法によって得られた非言語行動データについて平均値や相関係数の観点から比較を行った。また、会話場面に対する第三者による印象評定との関連についても検討した。分析の結果、アダプターと発話については動画・音声解析ソフトによる抽出法の方が平均値は小さくなることが明らかになった。ジェスチャーについては 2 つの抽出法間に有意な差はみられなかった。また、異なる抽出法を用いて得られた同じ非言語行動はそれぞれ有意な正の相関を示した。第三者による印象評定との関連については、イベントレコーダーにより抽出された非言語行動の方が高い説明率を示したものの、動画・音声解析ソフトによる抽出も有意な説明率をもつことが明らかになった。

キーワード: 非言語行動、動画解析ソフト、音声解析ソフト、イベントレコーダー、二者間会話.

問題

非言語行動は対人コミュニケーションにおいて重要な要因である（大坊, 1998）。親密さの表出(e.g., 大坊, 1990; 和田, 1996)だけにとどまらず、話者に対する対人印象(磯・木村・桜木・大坊, 2003)、社会的スキルや関係継続動機との関連(木村・磯・大坊, 2004)、会話満足度との関連(木村・大坊・余語, 2010)などが報告されており、その他にも枚挙に暇がない。また、対人コミュニケーション分野にとどまらず、近年の認知科学では荒川・西尾(2007)が非言語行動(論文中では、身振りとして言及されている)の抑制が課題のパフォーマンスを阻害するなど、以前から身体的動作と思考との関連を指摘する研究は散見される(e.g., McNeill, 1992)。様々な分野を越えて非言語行動は検討対象となっているのである。こうした研究は一般的に非言語行動の抽出や操作を行い、目的とする変数との関連を検討することにより行われる。

しかし、非言語行動の抽出にはいくつかの困難が伴う。まず、抽出に関する熟練度の問題である。一般的な手法として、非言語行動を抽出する際にはあらかじめ設定した定義に従って複数名のコーダーによる抽出が行われる。そして、各コーダー間の一致率や相関係数、α係数やκ係数などを算出することで信頼性を検証し、十分な信頼性が得られたと判断された段階で各コーダーの平均値を分析対象に用いることが多い(例えば、木村他, 2010)。このとき、高い信頼性を備えたデータを得るためにはコーダーの高い熟練度が必要とされる。また、抽出対象となる非言語行動が微細であればあるほど信頼性も高くなりにくく、コーダーの熟練度も相応に高い水準が要求されることになる。学術論文においては信頼性の高いデータを得ることが前提となるため、簡便さを求めて熟練度の低いコーダーを用いることはできない。つまり、非言語行動を抽出するためには必然的にコーダーの熟練度を高めざるを得なくなる。こうしたコーダーの熟練度についての問題は、非言語行動に関する研究を妨げる要因となっていると指摘できるであろう。

また、抽出にかかる時間と労力が多大になる点も挙げられよう。ビデオカメラの導入(e.g., 大坊, 1982)以降、時系列に沿って映像を視聴しながら非言語行動を抽出するのが一般的となった。これは、抽出対象となる映像を等倍速で再生することを意味し、抽出にかかる時間が映像の所要時間に依存することを意味している。また、いくらコーダーが熟練したとしても、一度に抽出できる非言語行動の種類には限度がある。例えば、視線とうなずきは同時に抽出できるかもしれない。なぜなら、視線とうなずきは顔に注意することで抽出できるため、注意の範囲が少なくて済むためである。さらに熟練度を高めれば、これらと同時に発話を抽出することも可能であろう。視覚的に視線とうなずきを抽出しながら、聴覚的に発話を抽出すればよいからである。しかし、1 人の話者についてこれら 3 つを同時に抽出することは可能であっても、もう一方の会話者の非言語行動も同時に抽出しようとするとコーダーの許容量を超えてしまうことが想像に難くない。視覚的な注意の範囲が増え、聴覚的にも 2 人の話者の発話を弁別する必要が出てくるためである。こうした場合、データの信頼性を保つためには 1 人ずつの話者に着目して抽出を行わざるを得ない。言い換えると、同じ会話映

像を複数回見て抽出しなくてはならないのである。こうした労力は、実験参加者数(映像数)が増えれば増えるほど、また、対象となる非言語行動が増えれば増えるほど加速度的に増加していく。さらに、連続して抽出に従事できる時間にも限りがある。何時間も集中し続けることはコーダーの負担になるだけでなく、疲労などによる集中力の低下を招くことでデータの信頼性を低める危険性がある。そのため、信頼性の高いデータを得るためには一度に従事する時間を短く区切らなければならず、結果としてコーダーを拘束する期間を増やさざるを得なくなる。これは結果的に、コーダーへの報酬など経済的な面でのコストが増加することも意味する。

こうした事情への打開策として、ごく短い時間の抽出を用いて話者の非言語行動の代替指標とする研究がある(e.g., Ambady, Koo, Rosenthal, & Winograd, 2002; Ambady & Rosenthal, 1992, 1993)。例えば、Ambady et al.(2002)では、理学療法士とクライアントの相互作用場面についての映像から20秒の映像を3つ抽出し、その中での療法士の非言語行動とクライアントの回復度について検討している。その結果、わずか 20 秒間の映像から抽出した療法士の非言語行動が、クライアントの 3 か月後における実際の回復度を予測することが明らかになった。そのほかでは、実際に非言語行動を抽出する代わりに、非言語行動の多少について評定者が評定するという方法を採用した研究がある(e.g., Guerrero, 1997)。ここでは、例えば笑顔や視線について 7 件法で評定を行うことで、実際の非言語行動の代替指標として分析を行っている。こうした研究はコーダーの熟練度に関する問題を回避できていないものの、必要となる時間や労力が大幅に削減できるという意味で節約的である。

しかし、これらの方法にも問題を指摘できる。ある一部分の非言語行動を抽出することは、少なくとも会話場面における時系列的な展開を度外視することになる。例えば、会話前に感情誘導を行った藤原・大坊(2010a)では、感情による非言語行動への影響は会話の比較的序盤に集中し、終盤になるにつれて感情の影響は弱くなることが明らかになっている。また、直前に行った 2 人会話時の非言語行動量が後続の 3 人会話時の非言語行動に与える影響についても、やはり序盤に集中しており終盤には影響が弱まっていくことが明らかになっている(藤原・大坊・前田・横山・前田・岸野・北村・林, 2010)。こうした時系列的な展開を追ううえでは、ある一部の時間帯のみを対象として抽出を行うことは不適当であるといえよう。また、評定者によって非言語行動の量を評定する方法についても、そのデータは非言語行動についての実測値ではなく、やはりあくまで評定値であり代替指標に過ぎないという点では限界であろう。

そこで本研究では、主にコーダーの熟練度の問題を回避するため、また、非言語行動の抽出にかかる時間や労力をある程度抑えるために新しい抽出法を導入する。具体的には、動画解析ソフトと音声解析ソフトを用いて非言語行動の抽出を行い、従来の抽出法との比較を試みることでその有用性や妥当性を検討する。なお、本研究では動画・音声解析ソフトを用いた抽出法には藤原・大坊(2010b)の方法を採用し、従来の抽出法としては、イベントレコーダーsigsaji2(荒川・鈴木, 2004)を用いて非言語行動の抽出を行った木村他(2010)の方法を採用することとした。

動画解析ソフトと音声解析ソフトを用いた非言語行動の抽出には 2 つの特徴が挙げられる。1 つは抽出に関するエラーの少なさである。藤原・大坊(2010b)では、動画内の両手や鼻先、両足の付け根の位置について 2 次元上で座標を測定し、手の位置を用いて非言語行動を抽出していた。すなわち、右手(左手)が右足(左足)の付け根よりも上にある状態で、鼻先よりも会話相手寄りにある場合(Figure1 内の①領域)をジェスチャー、自分寄りにある場合(Figure1 内の②領域)をアダプターとして抽出している。手が足の付け根よりも下にある状態は基本姿勢として抽出対象に含めていない。また、発話の抽出については独自に作成した音声解析ソフトを用いることで、入力された音声に対して閾値を設定し 1-0 データとして発話の有無を抽出していた。このとき、動画解析ソフトによる座標の測定や音声解析ソフトによる音声の識別は自動的・機械的に行われるため、見逃しや誤反応といった抽出上のエラーが起こる可能性はほとんどない [2]。従来のように映像を視聴しながらイベントレコーダーを用いて非言語行動を抽出する際には、どうしてもデータ内に見逃しや誤反応が入り込む危険性があった。ある程度はコーダーの熟練度を高めることにより回避できたものの、等倍速で再生される会話場面を対象とするため、全くエラーのない抽出を目指すことには限界がある。本研究で新たに採用する動画・音声解析ソフトによる抽出法ではエラーの少ないデータが得られることが特徴であり、かつ、それがコーダーの熟練度に関係なく達成されるという点に大きな意義があるといえよう。

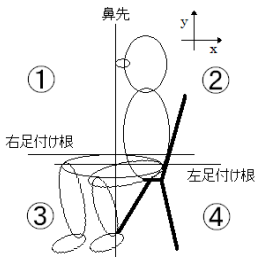

Figure1 非言語行動の抽出パターン(藤原・大坊, 2010b)

2 つ目の特徴としては、複数人のコーダーを必要としない点である。動画・音声解析ソフトによる抽出は自動的・機械的に行われる。これは、コーダーの熟練度に関係なく客観的なデータが得られることに加えて、誰が抽出しても同じデータが得られることを意味している。すなわち、自動的・機械的な抽出法そのものに信頼性が担保されるため、非言語行動を複数人が抽出する必要がないのである。言い換えると、ある 1 人が抽出を行うだけで信頼性の高いデータ得られるのである。これにより、同じ映像について複数人が抽出を担当しなければならないという負担は軽減され、また、研究者側にとってもコーダーを雇用する際の経済的コストを削減することができるのである。

本研究では、それぞれの抽出法によって得られた非言語データについて平均値や相関係数の観点から比較を行う。これにより、2 つの抽出法の間にどのような違いがあり、また、どの程度の類似がみられるのかを検討する。加えて、会話場面に対する第三者による印象評定との関連について検討する。これにより、第三者による印象評定という外的変数との対応関係について、どちらの抽出法がより説明力を持つのかについて比較する。なお、異なる 2 種類の抽出法を同じ会話場面に適用し、比較を試みた先行研究がないことから、本研究では仮説を設けずに検討することとする。

方法

本研究は 1)会話実験、2)非言語行動の抽出、3)第三者による印象評定から構成された。

会話実験

会話実験の参加者は関西地区の男女大学生 86 名で、初対面の同性同士によって会話が行われた[3]。実験の実施時期は 2009 年 6 月下旬～7 月下旬であった。そのうち本研究で分析対象とされたのは、動画や音声が鮮明であった 35 組 70 名(M = 18.65 歳, SD = 0.70, 男性 42 名, 女性 28 名)の会話場面であった。各参加者はお互いがよく知り合えるように約 6 分間の会話を求められた。会話の開始と終了の合図にはベルを鳴らした。参加者の座る椅子は回転式ではなく固定式であり、座席間の距離は 80cm であった。会話場面は 250cm の距離で真横に置いたデジタルビデオカメラ(HDR-SR12; SONY)を用いて撮影した。このとき、話者の発話を測定するための指向性マイク(ATM75; audio-technica)をそれぞれの参加者に装着してもらった。それぞれモノラルで録音し、2 本のモノラル端子を結合してカメラにつなぐことで 2ch の音源とした。これは、音声解析ソフト内で話者の発話を識別するためであった。

非言語行動の抽出

本研究では、対象とする非言語行動を「ジェスチャー」、「アダプター」、「発話」の 3 種類とした。これらを対象とした理由は、着座状態での会話場面について 1 つの固定カメラで真横から撮影した会話映像を用いる場合に抽出しやすい行動であったためである。非言語行動の抽出には二種類の方法を用いた。1 つは動画・音声解析ソフトを用いた観察による抽出法であり、もう 1 つはイベントレコーダーを用いた抽出法であった。

動画・音声解析ソフトを用いた抽出 ジェスチャーとアダプターについては、藤原・大坊(2010b)に基づき動画解析ソフトの Dipp-Motion XD Ver.3.20-2(ディテクト)を用いて第一著者が抽出した。具体的には、まず 1 秒に 30 コマが記録されている動画ファイル(avi ファイル)を 14 コマ飛ばしに静止画ファイルとして記録し直すことで時間分解能を 0.5 秒に対応させた。その静止画ファイルを用いて、右手と左手にポイントを付すことで手の位置を同定した。次に、相対参照点として鼻先と右足付け根、左足付け根にポイントを付した。これら両手と相対参照点の座標値(位置関係)を用いて、右手なら右足付け根、左手なら左足付け根よりも上の位置に手があり、かつ鼻先よりも相手側である場合を「ジェスチャー」として操作的に定義した。また、鼻先よりも自分側である場合を「アダプター」として操作的に定義した。これらの定義は、具体的な動作について言及していない点で木村他(2010)とは異なるものの、当該動作を行うために手がどの位置にあるのかという位置情報を用いて定義を行っている点に特徴があるといえる。

また、発話については参加者に付けてもらったマイクによって録音した音声に基づき、音声解析ソフトの Talk analyzer Ver.1-2-4 [4]を用いて発話量を算出した。それぞれ 0.5 秒単位で 1-0 データとし、総時間を分析に用いた。

イベントレコーダーを用いた観察による抽出 イベントレコーダーには sigsaji2(荒川・鈴木, 2004)を用いた。これは Microsoft 社製 Visual Basic で作成されており、PC 上で操作できる。時間分解能は 0.5 秒となっており、特定のキーを押すと 1 を記録し、キーを押していない場合に 0 を記録するように作成されている。これを用いることで、対象とする行動が生起していたか否かを 1-0 形式で記録した。各非言語行動の操作的定義としては木村他(2010)に依拠した。非言語行動を抽出する際には会話開始のベルと同時にイベントレコーダーを開始させ、会話場面を見ながら対象となる行動が生起した時に定めておいたキーを押すようにした。対象となる行動が持続している場合には、持続しているだけキーを押し続けた。こうすることで、非言語行動が生起した総時間と総頻度を定量化した。

コーダーは、第一著者と訓練を受けた男子大学生3名の計4名(M= 22.50 歳, SD= 2.38)であった。各コーダーの負担を軽減する目的で、4 名が全ての会話場面について抽出を担当するようにはしなかった。まず 4 名全員で集まり、それぞれ抽出の際の操作的定義を確認した。そのうえで会話場面をランダムに 1 つ選び出し、実際に 4 名で抽出を行った。各コーダーの総時間と総頻度を比べ、抽出における細かい確認事項(膝の上で行ったジェスチャーはカウントする等)とそれぞれのコーダーの特徴(発話の短い切れ目の際にキーを押し続けていた等)を指摘し合った。その後、1 つの会話映像について必ず 2 人のコーダーが担当するようにそれぞれ割り当てを決めた。コーダーの組み合わせについては、${}_4C_2$ = 6 種類の組み合わせを均等に配分した。

第三者による印象評定

本研究では、抽出した非言語行動との関連をみる目的で、第三者による会話場面への印象評定も実施した。評定者は男女大学生4名(M= 22.25歳, SD= 0.50, 男性2名, 女性2名)であった。評定項目は1)好意的であった、2)真面目であった、3)積極的であった、の 3 つであり、7 件法で評定を行った。これら 3 項目を用いたのは、林(1978)による対人認知の 3 次元(個人的親しみやすさ、社会的望ましさ、力本性)に基づく印象評定を行うためであった。

結果

本研究の目的を達成するために、まず、異なる 2 つの抽出法によって得られた非言語行動のデータについて平均値と相関係数を検討する。そのうえで、第三者によって評定された印象データとの関連について検討する。

抽出法間の平均値と相関係数

あらかじめ、イベントレコーダーを用いて抽出した非言語行動について、測定の信頼性を確認するためにコーダー2名ごとに *Pearson* の積率相関係数を算出した。その結果、全ての非言語行動で $r \geqq .82$ となったため、2名のコーダーの平均値を分析に用いることとした[5]。

動画・音声解析ソフトにより抽出した非言語行動とイベントレコーダーによって抽出した非言語行動との間に量的差異が認められるかについて検討するために、対応のない t 検定を行った。なお、いずれの非言語行動にも分布の偏りがみられたため、対数変換を施した値を分析に用いた。その結果、ジェスチャーには有意な差はみられなかったものの($t(138) = 0.10$, ns)、アダプターと発話についてはいずれもイベントレコーダーの方が有意に高い値となった($ts(138) > 3.37$, $ps < .01$)。この結果は、動画・音声解析ソフトによる抽出は人による観察に比べて判断の閾値が高くなることを示しているといえる。

次に、変数間の相関関係について着目する。異なる抽出法間の同じ非言語行動間(例えば、動画・音声解析ソフトによって抽出されたジェスチャーとイベントレコーダーによって抽出されたジェスチャー)には有意な相関関係がみられた。ジェスチャーで最も強い正の相関関係がみられ、アダプターと発話についても同様に有意な正の相関関係がみられたことから、動画・音声解析ソフトによる抽出は従来の抽出法との関連が認められたといえるであろう。また、同じ抽出法内における異なる非言語行動間の相関関係に着目すると、いずれの抽出法においてもジェスチャーはアダプターと発話の両方と有意な正の相関関係を示し、アダプターと発話には有意な相関関係がみられなかった。これについても2つの抽出法の関連を示しているといえよう。しかしながら、相関係数自体には違いがみられ、動画・音声解析ソフトではジェスチャーとアダプターの方が高い相関係数を示したのに対し、イベントレコーダーではジェスチャーと発話の相関係数の方が高いことから、抽出法間の差異も示唆された。

第三者評価との関連

まず、評定の信頼性を確認するために各項目について4名の評定値を対象として *Cronbach* の α 係数を算出

Table 1 各非言語行動における平均値(標準偏差)と相関係数

		M	SD	2	3	4	5	6
動画・音声解析ソフト	1 ジェスチャー	1.48	0.50	.65 **	.44 **	.81 **	.56 *	.54 **
	2 アダプター	1.26	0.68		.20	.37 **	.71 **	.35 **
	3 発話	2.21	0.13			.47 **	.03	.60 **
イベントレコーダー	4 ジェスチャー	1.49	0.45				.30 *	.60 **
	5 アダプター	1.61	0.57					.22
	6 発話	2.29	0.10					

$*p < .05$, $**p < .01$

注)分析には対数変換した値を用いた。

した。その結果、好意的であった: $\alpha = .70$、真面目であった: $\alpha = .45$、積極的であった: $\alpha = .83$ となった。そこで α 係数が極端に低かった項目については分析から除外し、好意的であった、積極的であった、の 2 項目を用いることとした[6]。

印象評定項目についての平均値はそれぞれ、好意的であった($M = 4.60$, $SD = 0.94$)、積極的であった($M = 4.78$, $SD = 0.99$)となった。また、2 変数間の相関係数は $r = .64$($p < .01$)であった。これらの第三者評定の印象と、それぞれの抽出法で得られた非言語データの値がどのように関連するのかについて検討を行う。そのために、各抽出法における非言語行動を説明変数、印象評定値を目的変数とした重回帰分析を行った。なお、非言語行動についてはいずれも分布の偏りがみられたため、対数変換を施した値を分析に用いた。各非言語行動の影響を検討するために変数選択は行わず、強制投入法を用いた。

動画・音声解析ソフトについて分析した結果、積極性の印象評定について非言語行動のジェスチャーと発話が有意な正の影響を与えることが明らかになった。また、好意度については非言語行動による有意な影響はみられなかった(Table 2)。

Table 2 動画・音声解析ソフトによって抽出された非言語行動が第三者評定に与える影響

	好意的であった	積極的であった
ジェスチャー	.23	.44 **
アダプター	.08	-.13
発話	.11	.30 **
Adjusted R^2	.08 *	.30 **

$^{*}p < .05$, $^{**}p < .01$

これに対して、イベントレコーダーでも積極性の印象評定について非言語行動のジェスチャーと発話が有意な正の影響を与えていた。しかしながら、動画・音声解析ソフトによる抽出とは異なり、好意度の評定にはジェスチャーが有意に正の影響を与えることが明らかになった。また、決定係数の値から、イベントレコーダーによって抽出された非言語行動の方が、印象評定の説明率が高いことが明らかになった(Table 3)。

Table 3 イベントレコーダーによって抽出された非言語行動が第三者評定に与える影響

	好意的であった	積極的であった
ジェスチャー	.39 **	.48 **
アダプター	-.02	-.16 †
発話	.16	.40 **
Adjusted R^2	.21 **	.56 **

$^{\dagger}p < .10$, $^{*}p < .05$, $^{**}p < .01$

考察

本研究の目的は、二者間会話場面における非言語行動の抽出法について新たな抽出法を採用し、その妥当性および有用性について検討することであった。そのために、従来のイベントレコーダーを用いた観察による抽出法と近年注目されつつある動画・音声解析ソフトを用いた抽出法により、ジェスチャー、アダプター、発話の 3 つの非言語行動を抽出した。

これら 2 つの抽出法によって得られた非言語行動データについて平均値の比較を行った結果、ジェスチャーについては 2 つの抽出法間に有意な差がみられなかったものの、アダプターと発話については動画・音声解析ソフトによる抽出法の平均値が小さくなることが明らかになった。アダプターについては、基本姿勢の捉え方が異なることに主な原因があると考えられる。動画解析ソフトを用いた抽出法では、足の付け根よりも手の位置が下にある場合(Figure1 内の③④領域)は全て一律に基本姿勢として捉えられた。一方で、イベントレコーダーを用いた抽出法では、本研究が依拠した木村他(2010)の定義に基づき足をさすったり膝の上で指を絡ませたりする動作もアダプターとして抽出された。こうした違いにより、動画解析ソフトによるアダプターが有意に少なく抽出されたと考えられる。ジェスチャーで有意な差がみられなかったことについては、相手に表出することが基本となるため膝下で行われることが少なかったことによると考えられる。発話については、録音上の問題が考えられる。つまり、マイクの取り付け方(マイクの距離や角度等)の都合でソフトウェアの閾値に十分反応できるほどの音量で音声を拾えていなかった可能性も考えられる。

それぞれの変数間の相関係数を検討した結果では、異なる抽出法を用いて得られた同じ非言語行動はそれぞれ有意な正の相関関係を示していた。ジェスチャーで

最も係数が高く、アダプターと発話についても比較的高い相関係数が得られたことから、抽出法間の十分な関連が示されているといえる。つまり、動画解析ソフトや音声解析ソフトを用いた抽出法が従来のイベントレコーダーによる抽出法の代替として実験研究に耐え得る指標である可能性を示したといえよう。しかしながら、これらの中では発話についての相関係数が他の2つの非言語行動に比べると劣るものであった。録音上の問題があったとするならば、さらに精度の高い録音環境の下で実験することで詳細な比較が可能になるであろう。そうした意味では、音声解析ソフトで抽出した発話が従来のイベントレコーダーで抽出した発話を代替する可能性については、本研究の知見のみで即断するよりも、今後さらに検討を重ねていくことで必要かもしれない。

さらに、同じ抽出法内における異なる非言語行動間の相関関係に着目すると、いずれの抽出法においてもジェスチャーはアダプターおよび発話と有意な正の相関関係を示した。しかしながら、相関係数の大小については傾向が異なっており、動画・音声解析ソフトによる抽出ではジェスチャーにおける相関係数はアダプターの方が高かったのに対して、イベントレコーダーによる抽出ではジェスチャーは発話と高い相関係数を示した。この結果については、アダプターの抽出法に主な原因があると考えられる。動画解析ソフトによる抽出では、右手(左手)が右足(左足)の付け根よりも上にある状態で、鼻先よりも会話相手寄りにある場合(Figure1 内の①領域)をジェスチャー、自分寄りにある場合(Figure1 内の②領域)をアダプターとして抽出していた。この定義に基づくと、もし髪を触ろうとして手を挙げたときに鼻先よりも相手寄りを手が通過していた場合には、先にジェスチャーが抽出されてその後にアダプターが抽出されることになる(手を膝上に戻すときには、さらにジェスチャーが抽出されることになる)。こうした動きにより、ジェスチャーとアダプターの相関係数が高くなった可能性が考えられる。

第三者による印象評定との関連については、いずれの抽出法においてもジェスチャーと発話が有意に積極さを説明していた。ジェスチャーが活発さと関連するとした先行研究(伊藤, 1991)を支持する結果となり、発話についても沈黙の少なさが活動性認知と関連することが示されていることから(小川, 2003)、先行研究を支持する結果であるといえよう。また、活発な会話行動は「直接性の高さ」として、関係の初期においては親密さの表れと捉えられる(大坊, 1990)。そのため、ジェスチャーが好意度に対して有意な影響を与えていたと考えられる。

なお、好意度と積極さのいずれにおいてもイベントレコーダーにより抽出された非言語行動の方が高い説明率を示していた。これについては、第三者評定もイベントレコーダーによる抽出も共に人の目による観察に依存した指標であることを踏まえると、容易に想定できる結果であったかもしれない。しかしながら、動画・音声解析ソフトによる抽出も有意な説明率をもつことが明らかになり、それ自体に十分な意義があるといえよう。なぜなら、本研究の結果は非言語行動の抽出をコーダーによる観察法という方法論的制限から解放するからである。特に、動画・音声解析ソフトによる抽出は方法論として熟練度の問題を回避できることを踏まえると、誰にでも非言語行動の抽出が可能になる点に大きな可能性があるといえよう。

限界と展望

本研究の知見を一般化する際にはいくつかの注意が必要である。まず、撮影環境の制限が挙げられる。本研究では、1 台のビデオカメラを用いて向かい合って座っている 2 人の話者を真横から撮影した。ジェスチャーやアダプターについて手の位置を用いて抽出できたのは、話者同士が正面に向かい合っていたという点に負うところが非常に大きい。また、カメラと話者の体の向きを固定するという意味で、回転しない椅子を用いることが重要であったことも看過できない点であろう。例えば、2 人が向かい合っておらず斜めに向き合っている場合には、手が鼻先よりも相手寄りにあるという定義が映像に反映されづらくなる。そのため、動画解析ソフトを用いた抽出法には工夫を施す必要性が生じる。このような場合には、複数台のビデオカメラを用意してキャリブレーションや同期を行うことで 3 次元データとして動画を扱う必要がある。今後の展望として、3 次元データを用いて手の位置の座標を特定し、その座標値を使って非言語行動として抽出することができるのかを検討していくことは、十分に意義があろうと考えられる。

また、第三者評定に用いた項目の少なさにも留意すべきである。本研究では林(1978)による対人認知の 3 次元に基づいた評定を行っており、基本的な評定にとどまっていた可能性は否定できない。また、評定者の負担を考慮したとはいえ、好意度と積極さをそれぞれ 1 項目で測定していたことにも議論の余地がある。そのため、動画・音声解析ソフトにより抽出した非言語行動が、応用的で状況に依存するような評価(例えば、課題解決場面における有能さ等)にどの程度説明力を持つのかについて

は今後の検討に譲らなければならない。とはいえ、こうした点については動画・音声解析ソフトによって抽出した非言語行動がある程度の説明力をもつという本研究の結果があってこそ生じた今後の課題であり、本研究で得られた知見の意義を損なうものではないと考えられる。

本研究は、非言語行動の抽出法について新たな方法を導入し、その妥当性と有用性を見出した。動画・音声解析ソフトを用いた抽出法は、会話場面を撮影する際の制約を考慮する必要がある一方で、コーダーの熟練度についての問題を回避できる点で有用である。本研究で得られた知見は、社会心理学のみにとどまらず、非言語行動について研究する様々な分野に資するものであろう。

引用文献

Ambady, N., Koo, J., Rosenthal, R., & Winograd, C. H. (2002). Physical therapists' nonverbal communication predicts geriatric patient's health outcomes. *Psychology and Aging*, **17**, 443-452.

Ambady, N., & Rosenthal, R. (1992). Thin slice of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, **111**, 256-274.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slice of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, **64**, 431-441.

荒川 歩・西尾 新 (2007). 身振りの相対的抑制が新奇な空間的問題解決に与える影響 認知心理学研究, **4**, 117-122.

荒川 歩・鈴木直人 (2004). しぐさと感情の関係の探索的研究 感情心理学研究, **10**, 56-64.

大坊郁夫 (1982). 二者間相互作用における発言と視線パターンの時系列的分析 実験社会心理学研究, **22**, 11-26.

大坊郁夫 (1990). 対人関係における親密さの表現—コミュニケーションに見る発展と崩壊— 心理学評論, **33**, 322-352.

大坊郁夫 (1998). しぐさのコミュニケーション サイエンス社

藤原 健・大坊郁夫 (2010a). 感情の社会性の再考 —社会的相互作用に着目した検討— 日本社会心理学会第 51 回大会発表論文集, 162-163.

藤原 健・大坊郁夫 (2010b). 覚醒度の異なるポジティブ感情の対人会話場面における機能 —会話満足度, および手の動きについての検討— 感情心理学研究, **17**, 180-188.

藤原 健・大坊郁夫・前田奈穂・横山ひとみ・前田貴司・岸野文郎・北村喜文・林 良彦 (2010). 複数回会話における会話特徴の変遷 —2 人会話がその後に行う 3 人会話に与える影響— 電子情報通信学会技術研究報告書, **110 (33)**, 79-84.

Guerrero, L. K. (1997). Nonverbal involvement across interactions with same-sex friends, opposite-sex friends, and romantic partners: Consistency or change? *Journal of Social and Personal Relationships*, **14**, 31-58.

林 文俊 (1978). 対人認知構造の基本次元についての一考察 名古屋大学教育学部紀要 (教育心理学科), **25**, 233-247.

磯 友輝子・木村昌紀・桜木亜希子・大坊郁夫 (2003). 発話中のうなずきが印象形成に与える影響 —3 者間会話場面における非言語行動の果たす役割— 電子情報通信学会技術研究報告書, **103(410)**, 31-36.

伊藤哲司 (1991). 対人相互作用場面におけるユニット的ノンバーバル行動の特性, 実験社会心理学研究, **31**, 85-93.

木村昌紀・大坊郁夫・余語真夫 (2010). 社会的スキルとしての対人コミュニケーション認知メカニズムの検討, 社会心理学研究, **26**, 13-24.

木村昌紀・磯 友輝子・大坊郁夫 (2004). 関係継続の予期が対人コミュニケーションに及ぼす影響 電子情報通信学会技術研究報告書, **104 (198)**, 1-6.

McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

小川一美 (2003). 二者間発話量の均衡が観察者が抱く会話者と会話に対する印象に及ぼす効果, 実験社会心理学研究, **43**, 63-74.

和田 実 (1996). 非言語的コミュニケーション —直接性からの検討— 心理学評論, **39**, 137-167.

註

1) 本研究の一部は日本社会心理学会第 51 回大会(2010 年)および HCS2011 年 5 月研究会、The 9th Biennial Conference of the Asian Association of Social Psychology(2011 年)において発表された。

2) ソフトウェアにおける精度上の問題から、自動追尾で完全に手や鼻先の座標が算出できないことがある。この場合には、静止画面に対して手動でポイントを付すことで座標値を算出することがある。この操作については機械的・自動的とはいえず、少なからず熟練度の問題が混入することは否めない。

3) 会話実験では会話に先立って感情誘導が行われ、各参加者の感情状態が測定された。また、会話後に会話や相手の評価が測定された。しかし、本研究ではこれらのデータは分析に用いていないことから詳述しない。

4) このソフトは、ビデオ機器からの音声出力を専用デバイス経由で PC 上に取り込み、0.5 秒単位で時系列に沿って発話の有無を 1-0 形式のデータとして記録する音声解析ソフトである。判定基準は PC 上に取り込まれた音声量(音の大きさ)とユーザーの指定した閾値であり、閾値を上回ったときに 1、下回ったときに 0 を記録する。

5) 各非言語指標についての *Pearson* の積率相関係数は以下の通りであった。便宜的に 4 名のコーダーをコーダーa, b, c, d とすると、ジェスチャーについてはコーダーa-b: r = .88, a-c: r = .88, a-d: r = .82, b-c: r = .99, b-d: r = .98, c-d: r = .96 であった。アダプターについてはコーダーa-b: r = .89, a-c: r = .97, a-d: r = .97, b-c: r = .98, b-d: r = .94, c-d: r = .87 であり、発話についてはコーダーa-b: r = .99, a-c: r = .95, a-d: r = .92, b-c: r = .96, b-d: r = .83, c-d: r = .93 であった(全て $p < .01$)。

6) 「好意的であった」の α = .70 も十分に高い値であるとは言えない。しかし、本研究の目的においては「積極的であった」の 1 項目を分析対象とするよりは第三者評定の印象との関連を広く検討する方が有用であると判断したため、2 項目を分析に用いることとした。

The extraction procedure of nonverbal behavior:

Comparison between dynamic-image and speech-signal analysis software and event recorder

Ken Fujiwara(*Graduate School of Human Science, Osaka University, Japan Society for the Promotion of Science*)
Ikuo Daibo(*Graduate School of Human Science, Osaka University*)

This study examined the reliability and validity of the new extraction procedure of nonverbal behavior from the context of a dyadic conversation. Nonverbal behavior (that is, hand gestures, adapters, and utterances) was extracted by two methods. One was the traditional observation that uses an event recorder. The other was a new method that uses dynamic-image and speech-signal analysis software. Interpersonal intimacy and activeness of each speaker were rated from the third person's point of view. Results showed that the amount of adapters and utterances with the event recorder were significantly higher than with the software. There was no difference between the results of the two methods regarding hand gestures. Same behavior that was extracted with different methods was positively correlated. Although the coefficient of determination was lower than the event recorder, nonverbal behavior extracted with the software could predict the rating of activeness and interpersonal intimacy from the third. These results showed that the reliability and validity of the extraction procedure using the software was verified.

Key Words: nonverbal behavior, dynamic-image analysis software, speech-signal analysis software, event recorder, social interaction.