

会話量子化器を用いた会話場面の記録

久保田 秀和^{†,††} 齊藤 憲^{††},
角 康之^{††} 西田 豊明^{††}

本論文は、個人の価値観に沿った形で会話を記録することを目的として、人が会話に参加しながら、会話中の特徴的な場面を映像として抽出するための装置である会話量子化器を提案する。会話量子は連続した会話の流れから切り出した意味的なまとまりを持つ最小の会話断片であり、会話量子化とはこの会話断片を持続的に蓄積することにより、元とは異なる会話の場において元の会話のエッセンスを再利用可能とする枠組みである。本論文では特に、会話量子の前提となる会話断片の獲得手法を検討するため、1 ボタン型のプロトタイプ版会話量子化器を開発する。ここでは、会話のどの部分が重要であったかは後になって気付かれがちであることに着目し、会話断片を切り出す時間区間の in 点と out 点をボタン押下数によって適及的に決定可能とする。また、操作者の話し手あるいは聞き手としての振舞いを妨げないことを考慮する。最後に、適及量決定手法に違いのある 4 種類の会話量子化器を用いて会議から会話断片を獲得する実験を実施し、操作記録とアンケート結果に基づいて、会話量子化器の設計と抽出された会話断片の持つ特徴について議論する。

Capturing Conversational Scene by Using Conversation Quantizer

HIDEKAZU KUBOTA,^{†,††} KEN SAITO,^{††} YASUYUKI SUMI^{††}
and TOYOAKI NISHIDA^{††}

This paper describes a conversation quantizer for capturing conversation by subjective judgement. A conversation quantum is extracted by identifying and encoding the segments of interactions in a conversational situation. A conversation quantization is another idea that is performed by reusing a large number of conversation quanta using a computer. We have prototyped a 1-button conversation quantizer to investigate into a method of segmenting conversation which is prerequisite for extracting conversation quanta. We pay attention to the late awareness of the important section in the conversation. Inpoint and outpoint of cutting out rewind to the past by pushing a button on the prototyped device. We also pay attention not to obstruct conversation while using such a device. We have conducted an experiment where conversation segments are extracted from the meeting conversation with four kinds of prototyped devices that have a difference in the rewinding amount. We discuss about design of conversation quantizer and nature of conversation segments from the result of manipulating records and questionnaire.

1. はじめに

会話は人の主たる知的活動であり、会話の記録は人が知識を構築するうえで大きな助けになると考えられる。その一方で、会話の場において自分が話し手あるいは聞き手として振る舞いながら同時にメモをとることは負担が大きい。映像や音声による記録でも重要と思われる部分のみを会話中に取り出して記録しておく

ことは難しい。近年では会話記録の自動化と利用性の向上を目指して、自然言語処理技術、映像処理技術、ユビキタスセンサ技術などを用いた会話の構造解析や自動編集がさかんに研究されている^{1)~3)}。会話を記録するにあたっては、だれにとっても重要さが分かる客観的な基準だけでなく、自分自身の価値観に沿った形で残したいことも頻りに経験される。しかし、計算機に支援された会話記録のプロセスに対して人の主観的な意図を反映させる方法はまだ十分に議論されていない。

本論文は、個人の価値観に沿った形で会話を記録することを目的として、人が会話に参加しながらかつ自分の意図どおりに、非言語情報も含んだ会話の要点を記録可能とするような会話量子化器 (Conversation

† 日本学術振興会特別研究員 PD
JSPS Research Fellow PD

†† 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University
現在、アクセント株式会社
Presently with Accenture

Quantizer) を提案する．まず，連続した会話の流れから切り出した意味的なまとまりを持つ会話断片を会話量子と呼ぶ．会話量子とはこの会話量子を持続的に蓄積することにより，元とは異なる会話の場において元の会話のエッセンスを再利用可能とする枠組みである⁴⁾．特に会話映像の中には興味深いコメントや言語化の困難な操作や技，発想を刺激するジェスチャなど会話参加者にとって知識の種となる場面が多数含まれていると考えられるため，本論文ではそのような有益な場面を会話映像から切り出した映像クリップを会話量子と考える．この映像ベースの会話量子を映像会話量子と呼ぶ．

会話量子化器とは人の主観的な意図に沿って映像会話量子を抽出するための装置である．本論文では会話量子の前提となる会話断片の獲得手法を検討するために，1ボタン型のプロトタイプ版会話量子化器を開発し，このプロトタイプを用いた実験を行うことにより，会話量子化器の設計と獲得された会話断片の持つ特徴について議論する．

2. 会話量子化

本論文では，会話を介して獲得される知識を個人の主観的・行動的な側面をとともう情報であるとしてとらえている．ここでは，会話の場において個人の主観的な判断によって会話の一部分を切り出すこと，また切り出した部分を本人が素材として再利用する行動を促進することが，知識の獲得をもたらすものと考えている．以上のプロセスを計算機的に実現するための枠組みを会話量子化と呼ぶ．会話量子化においては，会話の流れから切り出した意味的なまとまりを持つ会話断片を共通の作法で計算機的に処理可能とすることにより，CSCW や情報可視化技術の適用を容易なものとし，持続的に増加する会話の知を人にとって編集しやすいものとする．

会話量子化のプロセスを図1に示す．近年，会話の場は実世界にとどまらず情報通信の世界ともつながっている．ここでは情報通信技術を用いて可視化された会話の場をランドスケープと呼ぶものとする．実世界には多数のオブジェクトがあり(図1「実世界オブジェクト」)，その一部は会話の話題となることもある(図1「話題オブジェクト」)．また情報通信の世界には多数のコンテンツがあり，その一部は話題コンテンツとなることもある．会話の場面(A)では，話題オブジェクトや話題コンテンツを参照しながら会話が行われている．量子化プロセス(図1(1))ではこの会話が量子化され，ランドスケープへ取り込まれる．構築プロセス

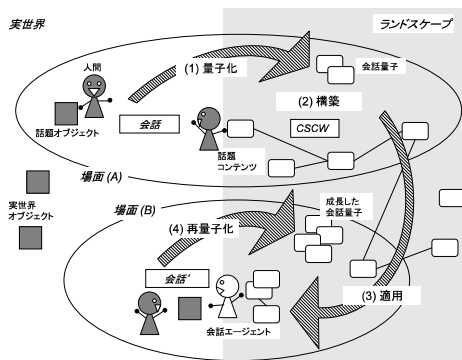


図1 会話量子化

Fig. 1 Conversation quantization.

ス(図1(2))では取り込まれた量子についてCSCW技術や可視化技術を用いた組織化が行われ，コンテンツが構築される．構築プロセスは量子を取り込んだ会話の場面または時間的空間的に別の場面において進められる．適用プロセス(図1(3))ではコンテンツとしてまとめられた会話量子が別の会話の場面(B)へ適用される．ここでコンテンツとは会話エージェントや会話ロボット用のコンテンツとして表現されることもある．人がプレゼンテーション資料として用いることもある．続いて再量子化プロセス(図1(4))において，場面(B)における会話は適用された会話量子と関連付けられて新たに量子化が行われることにより，再びランドスケープへ取り込まれる．以下，(2)構築プロセス，(3)適用プロセス，(4)再量子化プロセスのサイクルを様々な会話の場で繰り返すことにより，全体の会話の知が発展する．

会話量子の表現手法については分身エージェント研究を通じて議論されてきた⁵⁾．分身エージェントとは実在する人間の代理として会話可能なエージェントであり，分身エージェント同士の会話を人が観察することによるコミュニティ知識共有支援や，分身エージェントと人が質問応答可能なコンテンツの実現について研究されている．ここではテキストや図版を用いた語録や口頭発表の記録を会話量子として蓄積し，分身エージェントの会話生成において再利用可能としているが，会話の非言語的な側面を十分に表現することはできなかった．

我々はこれまでに会話映像から手作業で映像会話量子を抽出し，シミュレーションを行った結果，映像会話量子もテキストや図版を用いた会話量子と同様に再利用可能であるという示唆を得ている⁵⁾．ここでは映像会話量子の利点として会話のリズムや雰囲気表現できる点が示唆された．映像会話量子にはほかにも人

のジェスチャやポスチャ、話題となる対象物とのインタラクションなどを含むことができる点も期待できる。本論文では特に量子化のプロセスについて詳述する。なかでも会話映像を量子化する前提となる会話断片の獲得手法と、実験によって獲得された会話断片に含まれる内容について詳細に議論する。

3. 会話量子化器

会話から映像会話量子を抽出するための手法として会話量子化器を提案する。会話量子化器は知球ルームを構成するモジュールの1つである。

3.1 知球ルーム

知球ルームは知球⁶⁾と呼ばれる球面型コンテンツアーカイブシステムをランドスケープとして用いることによって会話量子化のプロセスを促進するシステムであり、知球のほかに会話を撮影するためのビデオカメラと会話量子化器および知球を表示するためのディスプレイから構成される(図2)。

知球ルームの典型的な利用法としては、ミーティングや作業のための場としての利用を考えている。従来のミーティングルームや作業スペースではその場で発生する会話を十分に再利用できなかったが、知球ルームでは会話量子化の枠組みに基づいて、プレゼンテーションシステム兼ホワイトボードとして知球を利用し、また会話量子化器を備えることによって、会話の再利用性を高めることができるものと想定している。知球のインターフェースはマウスと大型ディスプレイ、あるいはタッチパネルディスプレイである。会話の様子はつねにビデオカメラを用いて撮影され、参加者による会話量子化器の操作に従って映像会話量子が抽出される。抽出された映像会話量子は知球上でサムネイル画像として可視化される。映像会話量子は抽出時に表示されていた知球上の発表資料と関連付けられて、その近傍に表示される。知球を用いることにより映像会話量子は会話の議事録として参加者間で共有される。

3.2 プロトタイプ版会話量子化器の特徴

意味的なまとまりを持つ映像会話量子を抽出する前段階として、現在進行中の会話から会話断片を切り出すための的確な in 点と out 点を与えることのできるプロトタイプ版会話量子化器を提案する。一般に会話を記録する際、あるシーンが重要であるかどうかはしばらく進んでみないと分からないことが多い。このため映像を切り出す in 点は時間が経った後から遡及的に指定できるようにする必要がある。また、会話の場において参加者が話し手あるいは聞き手として自然に振る舞うためには、記録や遡及のための操作が会話を

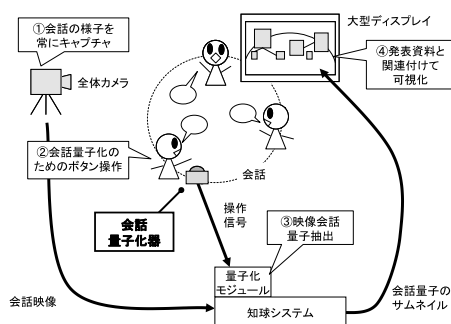


図2 知球ルーム

Fig. 2 Sustainable knowledge globe room.

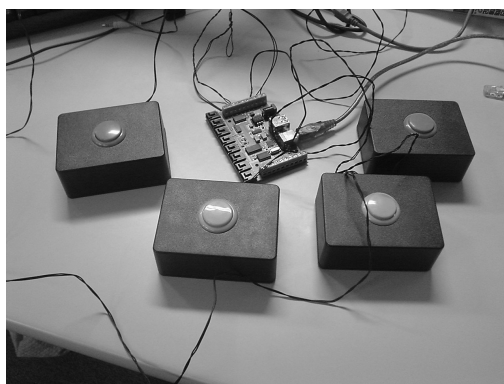


図3 1ボタン型会話量子化器

Fig. 3 1-button conversation quantizer.

阻害しないものでなくてはならない。我々は以上の要件を満たすインターフェースとして1ボタン型のシステムを採用する(図3)。会話を阻害しないという観点では、掴んで回すダイヤル式や複数ボタンを持つインターフェースは手間であり、1ボタンを叩く仕組みが良いと考えた。また、1ボタンあれば押下回数によって遡及量を十分に指定できるものと予想した。

3.3 予備調査

会話断片を切り出すための in 点 out 点を与える仕組みを検討するため、予備調査として、会話参加者がボタンを押したままにした区間の会話映像を切り出すことのできる装置を作成し、会議から会話断片を獲得した。会議は3名で実施し、1名のプレゼンタが自分の興味があるテーマについてスライドを用いて発表し、他の参加者と自由に議論した。会議ではプレゼンタのスライド発表に対して、残り2名の参加者がいつでも自由に質問やコメントなどを発言できるものとした。ここでは会話量子化器を1台用いて、参加者のうちプレゼンタでない1名が量子化器の操作者となった。操作者は会議の途中で残りもう1名の参加者と交替するものとした。

ボタンには切り出しの in 点を遡及するための機能も与えた．通常，in 点はボタンを押し始めた時点であり，ボタンを離れた時点が out 点となる．このとき押し始めの時点で余分に連続押下すると，1 度押下するごとに in 点が 5 秒遡るようにした．たとえばボタンを 3 度余分に押下した後で押したままにすると，初めに押下した時点から 15 秒遡った点が in 点となる．この装置を用いて会話参加者が会話中に切り出しを意図した時点と，切り出したいと思う範囲のパターンを調査したところ，参加者が切り出したいと思う範囲には図 4 に示すような 3 つのパターンが見られた．図 4 は A, B, C の 3 者による会話において，発話のタイミングと発話内容，および参加者のうち 1 名が会話断片を切り出そうと意図した時点とその範囲を Anvil⁷⁾ を用いて分析し，注釈を加えたものである．図 4 における「意図した時点」とは，操作者が会話断片を切り出すためにボタンを押下した時点を目指す．また，そのボタン押下操作によって記録される会話の範囲を「記録したい範囲」として示した (I) の例では，A と B がキュービズムの語源について会話した後に，話を聞いていた C が「へえ…」と感嘆の声をあげてボタンを押している．C はボタンを連続押下した後，すぐに離れたため，およそ会話の始まった過去の時点から C が切り出しを意図した時点までが記録されている (II) の例では，B が A の答えに感心して，自分と A との会話に対してボタンを押している．B はボタンを連続押下した後，押し続けたため，およそ会話の始まった過去の時点から今後続いてゆく会話が記録の範囲となっ

ている (III) の例では，C が A と B へ向けて質問をした時点から自分でボタンを押し続けており，連続押下はない．ここでは自分の質問とそれに対する答えの切り出しが期待されているものと思われる．

また，どこまで話がひと段落しているかということも少し時間を置かなければ判断できないものであり，in 点だけでなく out 点の指定にも遡及が必要であることが分かった．

4. 会話断片獲得実験

会話量子化器の設計と会話断片の持つ特徴について検討するために実験を行う．実験では 3.3 節の予備調査に基づき，遡及量決定手法に違いのある 4 種類のプロトタイプ版会話量子化器を用いて会議から会話断片を獲得する．

4.1 実験の手順

実験実施者は筆者のうちの 1 名であり，実験参加者は 20 代前半の男女 8 名 (男性 6 名，女性 2 名) である．実験では，参加者 3 名による 30 分程度の会議を 8 回行う．会議の目的は，1 名のプレゼンタが自分の興味があるテーマについて，スライドを用いて発表することによって他の参加者と自由に議論することである．ここでプレゼンタは必ず実験実施者が務める．テーマは偏らぬよう「医療問題」「近代美術」「京都大学」「お国自慢」「心理学」といった様々なものを用意して各会議へ割り当てた (表 2 を参照)．会議ではプレゼンタのスライド発表に対して，残り 2 名の参加者がいつでも自由に質問やコメントなどを発言できるものとした．会議では 1 台の会話量子化器を用いて，参加者のうちプレゼンタでない 1 名が量子化器の操作者となった．操作者には操作法を伝えたほか「議事録をとるつもりでボタンを押してください」と教示した．残り 1 名は特に制約のない立場で会議に参加した．また，会議は 1 台のカメラで全体が収まるように撮影した．

会議は遡及量決定手法 (ボタン押下パターンと 1 押下あたりの遡及秒数) が異なる 4 つの仕様 (表 1) について，それぞれ 2 回ずつ実施する．押下パターンとしては押下保持あり (図 5 (a)) と押下保持なし (図 5 (b))

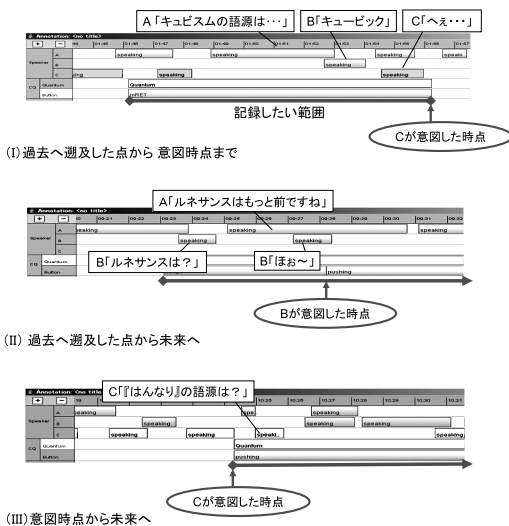


図 4 会話断片の切り出しを意図した時点と範囲

Fig. 4 Intended moment and range for conversation segmentation.

表 1 実験で用いるプロトタイプ版会話量子化器の仕様

Table 1 Specifications of the prototyped conversation quantizer in the experiment.

仕様	in 点遡及 単位 (秒)	out 点遡及 単位 (秒)	ボタン押下 パターン
1	3	3	保持あり
2	5	5	
3	10	10	
4	5	0	保持なし

表 2 会話断片獲得実験結果
Table 2 Results of the experiment for conversation segmentation.

仕様	テーマ	操作者	会議時間 (秒)	断片数 (個)	断片合計 (秒)	断片の 平均長(秒)	断片合計 /会議時間(%)
1	近代美術 医療問題	A	1,680	16	680	42.5	40.5
		B	1,200	12	466	38.8	38.8
2	医療問題 お国自慢	C	1,860	10	834	83.4	44.8
		D	1,800	15	645	43.0	35.8
3	お国自慢 近代美術	E	1,740	21	1,073	51.1	61.7
		F	1,620	11	786	71.5	48.5
4	京都大学 心理学	G	1,800	22	360	16.4	20.0
		H	1,980	20	529	26.5	26.7
平均			1,710.0	15.9	671.6	42.3	39.3

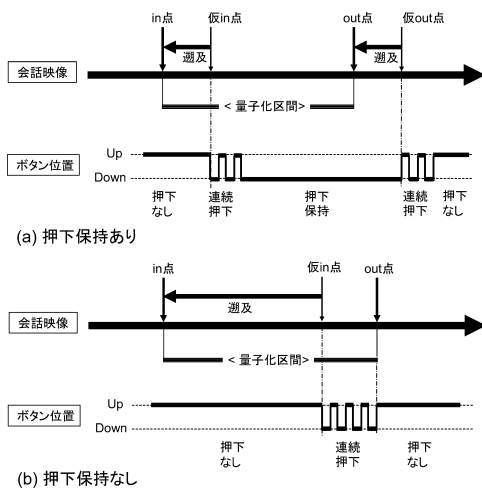


図 5 プロトタイプ版会話量子化器の押下パターン

Fig. 5 Push pattern of prototyped conversation quantizer.

の 2 つを用いる。

押下保持ありのパターンでは、まず初めのボタン押下点を仮 in 点とし、連続押下数に in 点遡及単位をかけた秒数だけ遡った点を最終的な in 点とする。また押下状態を保持した後、ボタンを離れた時点を決して仮 out 点とし、連続押下数に out 点遡及単位をかけた秒数だけ遡った点を最終的な out 点とする。in 点 out 点ともに連続押下は行わなくてもかまわない。ここで押下状態を保持する理由は、記録中であることを押下保持によって明示的にしないと、in 点を決定するための押下か out 点を決定するための押下かの判断が操作者にとって難しいと考えたためである。

またここでの遡及単位は、参加者が 10 秒程度経過後でボタン押下の判断を行うという予備調査で得た仮説に基づいている。およそ 10 秒遡るためのボタン押下回数が 3 回、2 回、1 回となるように、遡及単位は 3 秒、5 秒、10 秒と設定された。

ボタン押下を保持することは操作者の話し手や聞き手としての振舞いを邪魔する可能性があるため、押下

保持なしのパターンも用意して比較することにした。これは予備調査における「(I) 過去へ遡及した点から意図した時点まで」の切り出し例(図 4 参照)に沿った方法であり、in 点のみ遡及可能で、out 点は連続押下が終了した時点となる。

4 つの仕様にはそれぞれ参加者を 2 名ずつ割り当て、1 つの仕様につきテーマを変えて 2 回の会議を行う。このとき、1 回目の会議では 2 名の参加者のうち 1 名を操作者とし、もう 1 名は特に制約のない立場で会議に参加する。2 回目の会議ではテーマを変えたとともに操作者も交替する。

会議終了後に操作者は、自分が記録したすべての会話断片の映像について動画再生ソフトを用いて見直すものとした。ここではまず、自分が意図した部分が記録されているかどうかを検証し、映像の前後で余分なところ、足りないところがあると判断した場合、切り出される前の連続映像と見比べて、何秒余分だったか、足りなかったかについて調査票に記入するよう教示した。

このとき、各会話断片について何を興味深く感じたかについても簡潔にコメントするものとした。ここでは、ボタン操作直後のコメントを得ることが理想的であるが、会話の流れが中断されるなどの問題があるため困難であり、会議終了直後のコメントをボタン操作直後のコメントの近似とした。近似であるため会議終了後に改めて会話断片を見たときの印象も含まれると予想されるが、30 分程度の短い会議の直後に操作者が自分の気持ちを思い出すことはおよそ可能であると仮定した。

また、アンケート調査を実施し、操作者は次の 6 つの問いに自由記述文で答えるものとする。

- Q1 意図した会話断片が獲得されていたかどうか。
- Q2 会議知識の獲得に役に立つかどうか。
- Q3 会話断片は何に使えるのか。
- Q4 ボタン操作は学習できるかどうか。

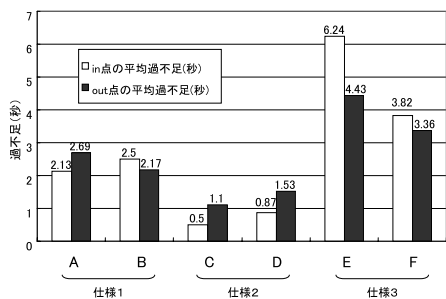


図 6 意図された範囲に対する平均過不足

Fig. 6 Average surplus/shortage to intended range.

Q5 ボタン操作学習のためのフィードバックは何が考えられるか。

Q6 その他、感想。

4.2 実験結果 1: 操作記録より

実験結果を表 2 に示す。各行につき、会議時間は会議を行った時間、断片数はその中で操作者が獲得した会話断片の数、断片合計は会話断片に含まれる映像の長さをすべて足し合わせた値、平均長は断片合計を断片数で割った値を示す。実験で得られた会話断片は合計 127 個、平均長は 42.3 秒であった。

4.2.1 遡及単位による比較

押下保持あり（仕様 1, 2, 3）の場合について、in 点 out 点の過不足の平均秒数を図 6 に示す。ここで平均秒数は超過分と不足分の絶対値を用いて求めた。図 6 からは仕様 3 のとき過不足の最も大きいことが分かる。また、各仕様における in 点遡及量の分布を箱ひげ図として図 7 に示した。図 7 では仕様 1（遡及単位は 3 秒）において遡及された最大の秒数が 6 秒であり、仕様 2（遡及単位は 5 秒）においては最大 10 秒であったことを読み取ることができる。つまり、仕様 1, 2 ではたかだか 10 秒しか遡及されておらず、仕様 3 でも中央値は 10 秒であることから、仕様 3 の 10 秒という遡及単位は大雑把にすぎたため、過不足が大きくなったと考えられる。以上のことは、人が会話量子化器を用いたときに遡及する秒数の平均的な値を知ったうえで、その秒数に合わせて遡及単位を設計すべきであるという示唆を与える。ただし、遡及量は会話のスタイルや話題によって変化し、たとえば、その場で考えながら話す場合と準備されたプレゼンテーションの場合とでは話速が異なるため、遡及量にも違いが生ずるものと予想される。

4.2.2 押下保持の有無による比較

表 2 において押下保持あり（仕様 1, 2, 3）押下保持なし（仕様 4）を比較すると、押下保持ありの場合は長い断片（平均長 54.1 秒）を少なく（平均数 14.2 個）、

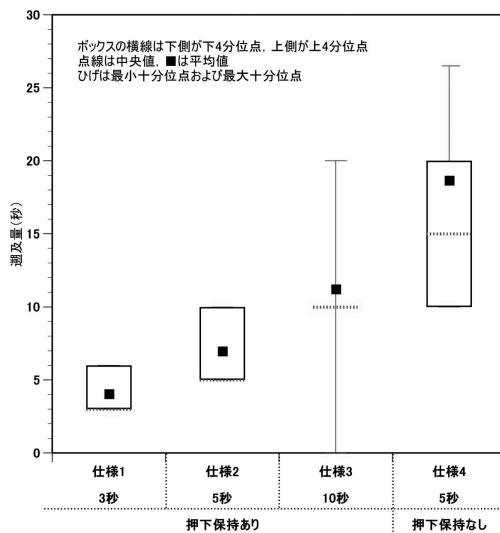


図 7 遡及量決定手法による比較

Fig. 7 Comparison by push pattern of conversation quantizer.

押下保持なしの場合は短い断片（平均長 21.4 秒）を多く（平均数 21 個）抽出するという傾向が見られた。

ここで図 5 に示したように、押下保持ありの場合は切り出し区間の初めに in 点の遡及操作を行う一方、押下保持なしの場合は切り出し区間の終わりで in 点の遡及操作を行う。このため in 点を決定する際、押下保持なしの場合は、切り出し区間の始まり（図中の「in 点」）がどれくらい過去であったかについて記憶を遡る必要のある時間（仮 in 点と in 点の幅）が長くなる。押下保持なしの場合の断片が短くなる理由は、現在の会話に参加しながら記憶を遡るのは困難であり、結果として遡るべき区間を短くしたものと考えられる。また同様に記憶できる長さの問題のためか、意図された範囲に対する遡及量の平均過不足は押下保持ありが 2.67 秒、押下保持なしが 9.62 秒であり、押下保持なしのほうが精度の低いことが確認された。

4.2.3 切り出しの境界とボタン操作のタイミング

操作者が記録を意図した in 点 out 点について、会議終了後本人から回答を得ている。この境界となる点が話者交替や無音区間、談話境界などに対応付け可能であるかを検討することによって、メディア処理やその他会話のセンシング技術、談話研究に対する本研究の位置づけを明らかにしたい。各参加者が記録を意図したすべての会話断片の in 点について、話者交替のあった場合、プレゼンタの用いたスライドの切替え

内容にかかわらずある時点で発声していた者を話者として、話者の変更が行われた時点の話者交替と呼ぶものとする。

表 3 会話断片の境界

Table 3 Boundaries of conversation segments.

参加者	話者交替 (個)	スライド (個)	無音区間 (個)	その他 (個)	すべての 断片数
A	2	1	2	11	16
B	2	1	4	6	12
C	3	1	1	6	10
D	9	1	2	6	15

のあった場合、無音区間(0.5秒以上の無音)のあった場合、いずれにも該当しない部分が in 点とされた場合(その他)の個数を表3に示す。ここでは4.2.1項4.2.2項より比較的操作者の意図とのずれが小さかった仕様1,2の場合について考えるものとする。また、話者交替、スライドの切替え、無音区間の3つについては2つの場合を同時に満たす断片も含むため、各場合の合計数は参加者の獲得したすべての断片数を超えることがある。

表3では参加者Dが話者交替を in 点とした会話断片を9個記録した場合を除くと、話者交替、スライドの切替え、無音区間よりもそれ以外の部分を in 点とする傾向が見られ、対応する映像では「僕の知り合いが言ってたんですけど」というような話の前置き後に in 点が置かれたり、フィラーや言い誤りを省くため in 点がある後に置かれるなど、発話中の操作者の記録したい箇所が文節単位で指定されていた。

スライドの切替えについて調査した理由は、本実験がスライドを用いた会話であるため、スライドの切替えが会話の大きな話題変更に相当するものと考えたためである。結果としてスライドが切り替わってすぐに in 点が置かれることはなかった。対応する映像では、スライドが切り替わって会話がしばらく進行した後、操作者が興味を持った箇所になってようやく in 点の置かれる様子が観察された。

以上よりプロトタイプ版会話量子化器を用いて切り出された会話断片は、話し手の発話の意図や準備された話題のまとめりよりは操作者側の意図や理解に沿った境界を持つものと考えられる。

また、in 点のボタンが押下されたタイミングを調査したところ、発話の途中で押下されることがほとんどだったが、「はー」「うん」など参加者の感嘆や相づちに反応したと思われる押下が各参加者に1,2件散見された。笑いに反応したのも同様の数であったが、参加者Dのみ15個中9個の断片が笑いに反応した点で押下されていた。表3の話者交替を in 点とする数でも参加者Dが突出しており、映像を確認すると参加者Dは場を盛り上げるようなテンポの良いやりとりを中心に記録していたことが分かった。

音声信号処理技術に基づいた話者交替や無音区間の検出は収録環境さえ整えられれば難しくないが、ボタン操作によって定められる境界のすべてをカバーすることはできないものと考えられる。境界の決定は操作者がどのような会話断片を記録したいかという意図や好みの影響を受け、また、操作者による談話境界の理解にも沿っていると予想される。

4.3 実験結果2: アンケートより

プロトタイプ版会話量子化器に関する実験後のアンケート結果を表4と表5に示す。Q1, Q2, Q4については肯定的なコメントと否定的なコメントに分類し表中に件数と例をあげた。なお1つの回答中に肯定的な部分と否定的な部分の両方を含む場合は両方の件数としてカウントしている。

Q1への回答は100%が肯定的なコメントであり、秒数のうえでは過不足があったにもかかわらず、会話断片には操作者の意図を反映できた部分が大きいと認識されている。また、コメント中には「ボタン操作をよく間違えた」などの操作の難しさに関する意見も含まれていた。参加者DとGは予備調査にも参加したため会話量子化器の操作経験が20分程度あるが、そのほかに特別なトレーニングは実施していない。Q4に対するDとG以外からのコメント「1回(知球ルームで)ミーティングを経験すればすぐに慣れると思う」(F)「学習できそう」(C)より、操作は学習できそうであるという示唆を得ているが、「会話の切りどころは学習が難しそう」(C)という意見も得ている。操作の学習可能性については今後、同じ操作者に対して継続的に利用実験することによって確認したい。Q5への回答では遡及量の加減を習得するため目の前に時計を置いたほうが良いという意見もあり、今後の実験の参考としたい。

Q2の設問文を実験後に再検討したところ、会議知識とは何かを厳密に定義しなかったため役立つかどうかに関する回答は参考にできないが、ノートとの比較検討を行う点について今後の実験計画の参考となるコメントを得た。

以下、4.3.1項では参加者のどのような意図が会話量子化器の操作に反映されたかを分類する。続いて4.3.2項では映像会話量子の役立つ場面についてヒントを得るため、Q3への回答についてまとめる。

4.3.1 会話断片の内容

本実験で得た127個の会話断片にはすべて、操作者による10字程度のコメントが付けられている。この

「だいたい」「そこそこ」以上を肯定的とらえた。

表 4 Q1, Q2, Q4 への回答
Table 4 Result of Q1, Q2, Q4.

	質問の内容	コメントの例 (カッコ内は回答した操作者名)
Q1	意図した会話断片が獲得されていたかどうか	肯定的コメント [8 件]: 「獲れていたが、ボタン操作をよく間違えた」(B) 「だいたいできていた。どう獲得するかの方針を決めるまでに時間がかかった」(F) 「多少の誤差はあっても大体は獲得されていた」(H) など
Q2	会議知識の獲得に役に立つかどうか	肯定的なコメント [8 件]: 「役立つと思う」(C, D, E, G), 「前もって方針を定めておけば役立つそう」(F) 「事実を切り出すには良い」(A) など 否定的なコメント [1 件]: 「ただし、話の流れから何か構造的なもしくは論理的なつながりを持ったものを切り出すにはまだノートのほうが良い」(A)
Q4	ボタン操作は学習できるかどうか	肯定的なコメント [8 件]: 「できると思う」(B, D, G, H), 「1回(知球ルームで)ミーティングを経験すればすぐに慣れると思う」(F) 「学習できそう」(C) など 否定的なコメント [1 件]: 「会話の切りどころは学習が難しそう」(C)

表 5 「アンケート Q3: 会話断片は何に使えそうか」への回答文
Table 5 Result of questionnaire 3: What is the use of conversation segments ?

	コメント (カッコ内は回答した操作者名)
(a)	「自分のプレゼン能力の向上」(A)
(b)	「スライドの 1 枚として、自分なりの意見の発露を示せる」(A)
(c)	「良いことを言うが印象派、非論理的な人の備忘録」(A)
(d)	「(雰囲気)を)かみしめられる」(A)
(e)	「自分の発言をもう 1 回思い出す。言いたいことをまとめられる。もっと面白く話せるかも」(B)
(f)	「ジェスチャの記録、会話参加者の同意(頷き)や疑問(首をかしげる etc)などの記録」(C)
(g)	「思い出メカとして使える」(D)
(h)	「その場の雰囲気をほかのメンバに伝えるのに適していると思う」(E)
(i)	「スライドあたりの映像が少なければ、知球に埋め込まれた形のままで議事録として機能すると思う。映像が多くなるとスライドのどこに映像が対応するかという情報も必要になりそう」(F)
(j)	「キャプチャした映像中の会話を文字情報として自動的に文字情報化できれば、議事録ができる」(G)
(k)	「自分用の議事録や他人との知識・経験の共有に使えそう(その場にいなかった人)」(G)
(l)	「議事録を作る人や、もう一度会議を振り返りたい人、ポイントを振り返るのに役立つ。会議に参加しなかった人)に知ってもらえるものとして使える」(H)

コメントと実際の会話断片の内容を照らし合わせた結果から、会話のどのような要素が注目されて切り出しが行われたかを分類した。分類とあてはまる断片の個数を以下に示す。なお 1 つの会話断片が複数に分類される要素を含むことがあるため、合計数 (161 個) は全会話断片数 (127 個) よりも大きくなっている。

スライドの説明 (83 個) 会議で利用したスライドについてプレゼンタが説明した内容に興味を持たれたりしている。お国自慢の話題であれば堺市は和包丁が特産であるといった内容を含む。

議論 (18 個) プレゼンタの説明に対して疑問が差し出されたり、自分の誤解を表明する様子に興味を持たれたりしている。写真中の人物がジャンヌダルクであるか否か、野洲市で大量の銅鐸が発掘された理由の検討といった内容を含む。

個人の体験 (51 個) 参加者の体験談を含む。男女や友だちの話など、だれと会って話してどんなことが起こったかなど具体的なエピソードである。

個人の身振り (7 個) 手や体の特徴的な動きを含む。営業において相手の潜在意識に訴える手の使い方、ふざけて身を乗り出すような動きの面白さなどが含まれている。

その他 (2 個) 意味のよく分からないものや以上にあてはまらないもの。

4.3.2 会話断片の用途

映像会話量子の役立つ場面についてヒントを得るため、アンケート Q3 (会話断片は何に使えそうか) への回答についてまとめる。回答はすべて表 5 にあげた。回答に含まれる要素は次のように分類することができる。

会話記録 会話の記録手法として会話断片を利用する。

回答 (c) では備忘録、回答 (i)(j)(k) では議事録としての利用があげられている。回答 (f) では特にジェスチャなど非言語的情報の記録に注目されている。また、回答 (l) では議事録以前のドラフトとして会話を振り返る際に会話断片が有用である

可能性について述べられている。

振り返り 会話を振り返って検討するために量子を利用する。回答 (a)(d)(e)(g)(l) からはただ記録しておくというだけでなく、積極的に振り返ることにより、自分の能力向上や思い出をかみしめることに利用したいという意図が感じられる。

再利用 会話断片を別の場所で再利用して会話内容を伝達するために利用する。回答 (b) では会話断片を発表資料の 1 つとして自分のプレゼンの中へ組み込みたい旨が述べられている。回答 (h)(k)(l) は知識共有の観点から、会話の場にはいないメンバーへの伝達手段として会話断片をとらえている。伝達内容については特に回答 (h) では会話の場の雰囲気伝えること、回答 (k) では知識や経験の伝達されることが期待されている。

雰囲気表現 会話断片を言語化の困難な雰囲気や印象というものを記録する手段として利用する。回答 (c)(d)(h) がこれにあたる。回答 (f) にあるような会話参加者の顔つきや首をかしげる様子も文字ベースの議事録には残されにくいという意味ではこの範疇である。

5. 議 論

本論文で提案したプロトタイプ版会話量子化器は、会話のうちどの区間が重要であったかは後になって気付かれがちであることに着目し、切り出しの in 点と out 点を遡及的に決定できることを設計上の特徴とする。4 章では会話量子化器の設計と会話断片の持つ特徴について検討するための実験を行った結果、4.3 節のアンケートに対する回答から、量子化器によって操作者の意図を反映した会話断片が抽出されていることが分かった。一方で、その操作法には難しさが残ると考えられていることも分かった。

また、会話断片には個人の身振りのような映像ならではの記録も含まれていることが分かった。会話断片の役立つ場面としては、記録として用いる場面、再利用可能な素材として用いる場面、映像ならではの身振りや雰囲気を伝える場面があるというヒントも得た。それぞれの場面で会話断片が会話量子として有効であるかどうかは今後の実験によって確認してゆきたい。

会話断片には会話の話題となる対象物を触ったり指差したりする場面が含まれると事前に予想していたが、該当する場面は見られなかった。実験中には、ボタンを押下保持していたために両手を用いたジェスチャの妨げられた様子が観察されており、その他の身体を使った動作にも押下保持が制約を与えていた可能性がある。

現在、押下保持が一定時間手を塞いでしまうという問題に対処するため、記録中であることを押下保持以外の方法で確認できる方法を提供するタッチパネルタイプの会話量子化器を開発中である。タッチパネルタイプの会話量子化器は、画面のタッチ回数により遡及のための操作が可能であると同時に、画面上に記録中であることを表示可能としている。また今後は、機器の組み立てや操作など作業が行われる場面での会話も適用対象としたい。4.2.2 項ではボタン押下保持なしの場合よりも保持ありのほうが高い精度を得られたが、作業の場面では特に保持ありによって手を塞がれることの不便が大きいと予想されるため、結果がどのように変化するか興味深い。

今回プロトタイプとして開発した 1 ボタン型の会話量子化器については、4.2 節の操作記録から、平均的な遡及量に合わせた遡及単位の設計と、記憶の限界の考慮が必要であるという示唆も得た。そのほか、アンケート設問 Q2 への回答として、話の流れから何か構造的なもしくは論理的なつながりを持ったものを切り出すにはまだノートのほうが良いというコメントも得た。従来のノートを用いた会話の記録と会話量子化器を用いた記録との比較も行いたいだが、会話量子化器で獲得される会話断片が映像を含む一方、ノートによる記録は絵や文字であるためメディアが異なり、またノートには空間的に記録を整理可能な側面があるため、そのままでは対照実験が難しい。会話断片についても知球上へ取り込むことにより、ノートに近い形で情報を空間的に配置したり、接続したりすることが可能となるため、今後、今回得た設計に関する知見を活かしつつ会話量子化器を知球へ組み込むことによって、ノートとの比較を可能な限り行ってゆきたい。

また、会話断片を用いて会議を振り返る場合と、収録した全会議映像を用いて振り返る場合との違いについて、映像を再生するために必要な時間の観点から比較する。表 2 により各参加者について会議時間に対する会話断片の合計時間の割合は 20% から 62% までの間であり、会話断片をすべて見る場合は会議映像をすべて見る場合と比べて最大でも 6 割の時間しかかからないことが分かる。よって、少なくとも量子化器を操作した本人については、会話量子化器を用いることによる事後の見直しにかかる時間の省力化について期待できると考えている。

今回の実験で得られた会話断片の内容は、量子化器の操作者の主観に基づくまとまりを持つ内容であったといえる。ここでは説明的な内容だけでなく個人の体験のようなカジュアルな内容も抽出されている。情報

抽出では一般に、重要箇所の候補を抽出する段階では計算機による自動処理が活躍するが、候補をあるサイズのまとまりとしてとらえ、編集するプロセスでは人間の思考が深く関与する。実験後に操作者が会話断片へ付けたコメントにも「新しい知識、なるほどと思ったこと」「勉強になった」というような文が含まれており、操作者にとって重要な箇所は本人の発見や知識量に依存していることが窺われる。今回提案した1ボタン型の会話量子器は素朴なシステムであるが、会話に対して意味のあるまとまりを簡単に与えられる点で効果的であり、計算機処理による情報抽出と組み合わせることによって、会話からの知識獲得を促進するものと考えられる。また、操作者の意図に沿って定められた切り出しの境界やボタンの押されたタイミングは、会話参加者による談話の理解を分析するうえでも興味深い。

今回の実験では参加者に対して自分が興味を持った部分を記録して欲しいという意味合いで「議事録をとるつもりでボタンを押してください」とのみ教示したが、だれのために議事録をとるか明らかでなく、厳密さに欠ける部分があった。ただし、実験参加者からは前述のような本人の発見や知識量に依存したコメントを得ていること、また実験では会話量子器のことを開発コードの「へえボタン」(テレビ番組で話題となった、自分が感心した事例に対して押すボタン型ツールの名称)と呼んだことがバイアスとなった可能性から、他人のためよりは自分のために切り出しを行った側面が大きかったと考えている。

今回の実験で獲得された会話断片が会話量子の要件をどの程度満たすのかについて、今後より深く検討する必要がある。意味的なまとまりの良いことが会話量子の要件の1つであるため、会話断片を切り出した本人以外による内容の理解度を調査したい。また、別の場面において再利用可能であることも会話量子の要件である。今回は会話断片の切り出しまでしか行わなかったが、今後、会話量子化のスパイラルに沿って別の場面での再利用を何度も行う形での実験を行うことにより、獲得される会話断片の再利用性について検討したい。

最後に、会話量子化器が記録のためだけでなく他人へメッセージを送るために利用された事例について述べる。実験では会話量子化器の操作者Dがプレゼンタの発話への返事代わりに「さん(プレゼンタの名)の優しさに」と発話しながらボタン押下する様子が観察された。このような場合、話し手にとっては逆に、聞き手のリアルタイムの反応をボタン操作から知

ることができるともいえる。一方で、「ボタンをいつ押しているか分からないと良い」という感想もアンケート回答に見られた。自分がボタンを押している様子を他人に見られたくない、あるいは他人がボタンを押している様子を見たくないという場合もまたあると考えられる。以上のように、会話量子化器が会話における新たなモダリティとして認識される様子は興味深く、今後分析を進めたい。

6. 関連研究

会話中の参加者の意図を情報抽出に利用する試みとしてはディスカッションマイニングプロジェクト²⁾があり、議論札と呼ばれる発言意図を示す電子札や発言に対する態度を表明するボタンを用いて参加者の意図を記録システムへ伝える手法が提案されている。ここではあらかじめ意図の割り当てられた数種類の札とボタンが用いられるが、本研究では参加者の多様な主観を反映できるよう、前もって種類を限ることなく意図を伝える手法を提案している。

マルチモーダル会話の記録と分析に関する研究は広く行われている。AMI⁸⁾は、会議ブラウジング技術の実現と会議コーパスの構築を目的として、マルチモーダル会話の記録と分析を進める国際的プロジェクトである。また、CHIL(Computers in the Human Interaction Loop)⁹⁾はオフィスや講義室における人間のマルチモーダルな活動を認識し、コンピュータを人間のインタラクションの輪の中に入り込ませる形で支援することを目的とする国際プロジェクトである。VACEプロジェクト¹⁰⁾では、米国防軍の戦闘ゲームセッションを対象として、多人数マルチモーダルインタラクションの記録と分析が進められている。また、Minohら³⁾は講義室におけるコミュニケーション活動について、角ら¹⁾はデモ展示会場における展示者と見学者のインタラクションについて記録・分析している。本研究も人のマルチモーダルな会話の記録を大きな目的とするが、話者の明示的な意図を重視する点において以上の研究とアプローチが異なる。

7. おわりに

本論文では人が会話に参加しながら、会話中の特徴的な場面を映像として抽出可能とする1ボタン型の会話量子化器を提案した。この会話量子化器は操作者の話し手あるいは聞き手としての振舞いを妨げない操作によって、会話断片を切り出す時間区間を指定できることを狙いとする。特に、会話のどの区間が重要であったかは後になって気付かれがちであることに着目

し、切り出しの in 点と out 点を遡及的に決定できることを特徴とする。会話量子化器の設計と会話断片の持つ特徴について検討するための実験として、遡及量決定手法に違いのある 4 種類のプロトタイプ版会話量子化器を用いて会議から会話断片を切り出した。実験の結果、量子化器によって操作者の意図を反映した会話断片が獲得されていることが分かった。会話断片には個人の身振りのような映像ならではの記録が含まれていることも分かった。また、1 ボタン型会話量子化器の設計については、利用の場面における平均的な遡及量に合わせた遡及単位の設計と、操作者の記憶の限界の考慮が必要であるという示唆を得た。

参 考 文 献

- 1) 角 康之, 伊藤禎宣, 松口哲也, シドニーフェルス, 間瀬健二: 協調的なインタラクションの記録と解釈, 情報処理学会論文誌, Vol.44, No.11, pp.2628-2637 (2003).
- 2) Nagao, K., Kaji, K., Yamamoto, D. and Tomobe, H.: Discussion Mining: Annotation-Based Knowledge Discovery from Real World Activities, *Proc. 5th Pacific-Rim Conference on Multimedia (PCM 2004)* (2004).
- 3) Minoh, M. and Nishiguchi, S.: Environmental Media - In the Case of Lecture Archiving System, *Proc. Int. Conf. Knowledge-Based Intelligent Information & Engineering Systems (KES2003)*, Vol.2, pp.1070-1076 (2003).
- 4) Nishida, T.: Conversation Quantization for Conversational Knowledge Process, *Special Invited Talk*, Bhalla, S. (Ed.): DNIS 2005, LNCS 3433, pp.15-33, Springer (2005).
- 5) Saito, K., Kubota, H., Sumi, Y. and Nishida, T.: Support for content creation using conversation quanta, *New Frontiers in Artificial Intelligence: Joint JSAI 2005 Workshop Post-Proceedings*, LNAI 4012, pp.29-40, Springer (2006).
- 6) 久保田秀和, 西田豊明, 角 康之: 等高線表現を用いたコンテンツアーカイブの可視化, 人工知能学会全国大会 (第 20 回) 論文集, 2F1-2 (2006).
- 7) Kipp, M.: Anvil - A Generic Annotation Tool for Multimodal Dialogue, *Proc. 7th European Conference on Speech Communication and Technology*, pp.1367-1370 (2001).
- 8) Carletta, J., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., Kadlec, J., Karaiskos, V., Kraaij, W., Kronenthal, M., Lathoud, G., Lincoln, M., Lisowska, A., McCowan, I., Post, W., Reidsma, D. and Wellner, P.: The AMI Meeting Corpus: A Pre-announcement, *Machine Learning for Multi-*

modal Interaction 2nd International Workshop, MLMI 2005, Revised Selected Papers, Renals, S., Bengio, S. (Eds.), *Lecture Notes in Computer Science*, Vol.3869, pp.28-39 (2006).

- 9) Waibel, A., Steusloff, H., Stiefelwagen, R. and the CHIL Project Consortium: CHIL: COMPUTERS IN THE HUMAN INTERACTION LOOP, *5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)* (2004).
- 10) McNeill, D.: Gesture Gaze and Ground, *Machine Learning for Multimodal Interaction 2nd International Workshop, MLMI 2005, Revised Selected Papers*, Renals, S., Bengio, S. (Eds.), *Lecture Notes in Computer Science*, Vol.3869, pp.1-14 (2006).

(平成 19 年 4 月 2 日受付)

(平成 19 年 9 月 3 日採録)



久保田秀和 (正会員)

1998 年京都大学工学部情報工学科卒業。2000 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。2004 年東京大学大学院工学系研究科博士課程修了。博士 (工学)。2006 年より日本学術振興会特別研究員 PD, 現在に至る。コンテンツ可視化, 会話情報学, インタラクションの研究に従事。



齊藤 憲 (正会員)

2004 年同志社大学工学部知識工学科卒業。2006 年京都大学大学院情報学研究科修了。修士 (情報学)。2006 年アクセンチュア (株) 入社, 現在に至る。



角 康之 (正会員)

1990 年早稲田大学理工学部電子通信学科卒業。1995 年東京大学大学院工学系研究科情報工学専攻修了。同年 (株) 国際電気通信基礎技術研究所 (ATR) 入所。2003 年より京都大学大学院情報学研究科助教授 (現在は准教授)。博士 (工学)。研究の興味は知識処理システムとヒューマンインタフェース。



西田 豊明 (正会員)

1977年京都大学工学部卒業．1979年同大学院修士課程修了．1993年奈良先端科学技術大学院大学教授，1999年東京大学大学院工学系研究科教授，2001年東京大学大学院情報理工学系研究科教授を経て，2004年4月京都大学大学院情報学研究科教授．会話情報学，社会知のデザインの研究に従事．著書・編著に、『インタラクションの理解とデザイン』，岩波書店（2005年），“Conversational Informatics: An Engineering Approach,” John Wiley & Sons, Inc.（2007年11月刊行予定）．日本学術会議連携会員．本学会理事，人工知能学会理事・編集委員長．
