

グループにおける意思決定への支援のための 発話群からの会話抽出方法の提案

Proposal of Method of Conversation Extraction from Utterances for Group Decision Support

河野 進†
Susumu Kono

相原 健郎‡
Kenro Aihara

1. はじめに

グループにおける意思決定では、必ずしもメンバー全員の意見が一致せず、決定事項に対して不満を感じるメンバーが生じることがある。この課題を解決するため、我々はグループダイナミクスや会話分析の先行研究をふまえて、会話を行っているグループの状態を推定し、そのグループ状態に応じて、適切な参考情報やアドバイスを提示するといった介入により、グループにおける意思決定を支援することを考えた。

グループダイナミクスの先行研究では、グループはメンバー間の関係、親密度、構造等の特性によって、特有の状態や挙動を示すことが明らかになっている[1-6]。また、会話分析関連の先行研究では、人の発話内容を認識しテキスト化する音声認識技術[33-34]、発話特徴量から感情を検出する技術や[24-26]、音声認識結果から発話意図を抽出する技術[35-38]が明らかになっている。これらを活用して、会話に参加している複数人の発話を分析した結果からグループ状態を推定し、そのグループ状態に応じた意思決定支援につなげたいと考えている。

例えば、グループにおける意思決定の場面において、対話エージェントによって、グループ状態に応じた適切な参考情報やアドバイスを与え、意思決定のための意見交換に活用してもらうことを想定している。特に意思決定の場面において、意見の表明ができていないなどといった弱い立場にある人に対して、意見を引き出すような支援を行うことによって、グループ全体の満足度を高めることを意図している。先行研究を活用したグループの意思決定支援方法の概要については2章で述べる。

こういったグループにおける意思決定への支援を、会話分析を通じて実現するためには、対象となる会話に含まれる発話を選別して分析することが重要であり、同時に複数の会話が行われる場合においても、発話群の中からそれぞれの会話に含まれる発話を抽出できることが必要となる。この会話抽出の提案方法については、3章で述べる。また、サンプル会話音声を使った予備実験の結果については、4章で述べる。

今後は、この会話抽出によって、会話内の各発話の内容と特徴量変化を紐づけて分析することにより、グループの状態を推定し、その状態に応じてグループにおける意思決定を支援する手法を具体化していきたい。

2. 関連研究とその活用

2.1 対象グループ

本稿では、「グループ」を共通のメンバー意識や目標を持ち、相互作用や影響を及ぼし合う一緒に活動する個人の集合体、「メンバー」をグループに属する個人、「話者」を声を発する個人と定義する。そして、利益拡大や社会貢献といったような高度な目標を迫及する企業内のタスクグループのような組織ではなく、特に、旅行やレジャーに自動車等で一緒に出かけ、楽しく時間を過ごすといったような、当該グループ内で決めた緩やかな目標を持つ2~10人程度のメンバーで構成されるグループを対象とした。

2.2 グループダイナミクスとその活用

グループは、そのメンバー間の関係、親密度、グループの構造（例：上下関係を持つ、フラットな関係を持つ）、タスクの種類（例：メンバーが興味のあるタスクかどうか、全メンバーの協力が必要なタスクかどうか）によって異なる状態や挙動を示し、その状態はメンバー間のやりとりや外部要因などによって変容を示すことが、グループダイナミクスの先行研究において明らかになっている[1-6]。また、グループをその親密度、タスクや社会的関係性に着目して分類する方法が示されている[1]。

グループが意思決定を行う際のプロセスにおいても、同様にメンバーの関係性やグループの構造、タスク等によって特有の傾向があることが明らかになっている[7-16]。

本稿では、このような当該グループにおいて長期的に築かれたメンバー間の関係、親密度、構造によって分類されるグループの種類を「グループ種別」と定義する。

2.3 グループにおける意思決定への支援

活動が継続的なグループに対しては、親密度などの対人関係に考慮して意思決定のプロセスに介入することが望ましいとされている[15]。そして、意思決定のための情報は、各メンバーに適切に認識できる形で共有されなければ、参照されない傾向があり[1,5-8]、他のメンバーの意見については、類似性、熟知性、近似性、報酬性が感じられると、その意見に妥当性を感じる傾向があるとされている[3,4,14]。

本稿では、2.1で述べたように旅行やレジャーに出かけるグループを対象とし、企業が経営に関わる方針を決めるような高度な意思決定ではなく、これから行く食事場所や立ち寄り地を、会話を通じて討議して決めるような意思決定を取り上げる。グループのメンバーがまだ決まっていない、あるいは変更の必要が生じている食事場所等を決める場面で、各メンバーが自分の意見を表明でき

† 総合研究大学院大学 SOKENDAI (The Graduate University for Advanced Studies)

‡ 国立情報学研究所 National Institute of Informatics

たかどうか、更にその意見が採択されたかどうかでグループの満足度を測り、複数回の決議の中で、より多くのメンバーの意見を決議へ反映し、グループの満足度を高めることを指向する。特に、我々は、討議の中で自分の意見を表明できていない弱い立場にあるメンバーから、その意見を引き出すような支援を行うことによって、グループ全体の満足度を高めることを想定している。

なお、本稿では、「グループ状態」を、グループの会話での発話の様子や心的状態、特にある時点での様子や状態と定義する。

上記のようなグループにおける意思決定への支援を実現するため、グループダイナミクスでの考え方に基づいて、会話を行っているグループの種別や状態を推定し、それらに応じて、適切な参考情報やアドバイスを提示する介入により、意思決定を支援する方法を検討した。

まず、「各メンバー間の関係」、「親密度」、「グループの構造」といった特性による次のような3つの種別に分類した場合、当該グループがいずれの種別にあてはまるかを自動的に判定する。その種別に応じ対話エージェント等を通じて適切な参考情報やアドバイスを与える。以下に、主なグループ種別ごとに適用する介入例を、図1にグループ対話への介入イメージを示す。

(1)フラットな関係で親密度が高いグループ（親友同士など）

このグループ種別では、メンバーは率直な意見交換ができると想定される。そのため、例えば、対話エージェントが、討議を深めるための参考情報を全員に与えて、意見交換を活発化させるように仕向けることが想定される。

(2)上下関係がある親密度が高いグループ（家族など）

このグループ種別では、年長のメンバーが他の各メンバーの嗜好を把握している場合がある。そのため、対話エージェントが年長者に参考情報を与え、その情報を参考にして、年長者が各メンバーとの間で意見を確認し合うようなことが想定される。

(3)上下関係がある親密度の低いグループ（職場の上司・部下など）

このグループ種別では、年長者がグループの意思決定を主導し、年少者は率直に意見を表明できない場合がある。そのため、グループの中で弱い立場にあるメンバーや孤立しているメンバーを見つけ出し、そのようなメンバーへ参考情報やアドバイスを与え、他のメンバーに対して自分の意見を表明できるように支援することが想定される（図1）。

2.4 発話分析とその活用

音声分析技術の先行研究において、グループ対話での同調傾向、親しさ、対話の盛り上がり等の状況は、音圧レベル (dB)、トーン (F_0 , Hz)、ピッチ (F_1 - F_4 , Hz)、発話速度 (mora/sec)、重なりといった発話特徴量に現れるとされている[17-23]。これらの発話特徴量は会話状況だけでなく、会話に参加する各話者の感情を推定することにも活用され[24-26]、音圧レベル等を使って算出されるメル周波数ケプストラム係数 (Mel-Frequency Cepstrum

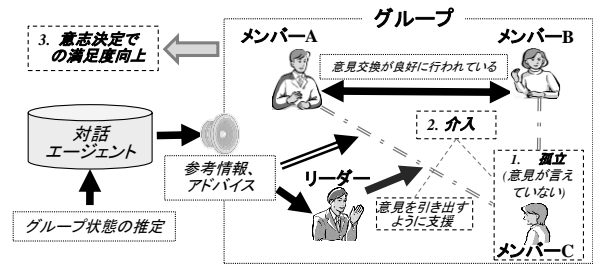


図1 グループ対話への介入イメージ

Coefficients) の分析から話者のテンション（緊迫感やストレス）が高い状態かどうかを判定できる[26]。

このような発話特徴量と会話における話者の状態や感情との関係性を活用して、抽出した発話特徴量の変移を発話内容と紐づけることによって、会話の進行と共に、各メンバーの状態がどのように変化しているのかを把握し、グループ全体の状態を推定する方法を検討した。

発話分析の前段の処理としては、まず、周囲の雑音や複数話者の発話が混在する音声信号から雑音を除去し、話者ごとの発話を分離して抽出（音源分離）することが必要となる。雑音除去や音源分離の技術では、周波数成分によって選別する手法や複数マイクを用いたマイクアレイによって到来方向別に分離する手法などがあり[27,28]、また、複数話者の発話の中から各話者の発話をベクトル量子化に基づくモデルとガウス混合分布モデル (Gaussian mixture model: GMM) を利用して識別する話者識別技術も活用できる[29]。

そして、音声信号から人が発声した発話区間を検出する技術では、音声信号の振幅によって検出する基本的な方法の他、GMMに基づき音声と非音声を分離する手法がある[30-32]。これらの技術によって、雑音の除去および音源（話者）の分離を行い、音声信号から個々の発話を抽出することが考えられる。

次に、音声認識技術を活用し、切り出された発話音声から、その発話内容を機械的に書き起こし、テキスト化することが必要となる。[33-34]。この音声認識によりテキスト化された発話内容のデータは、この後の処理となる発話意図抽出に活用することを想定している。

なお、本稿において、「音声」を音や声の音響信号、「会話」を複数の話者が発声して行う対話、「発話」を連続して行われている会話の中で前後の間隔によって区切られる1話者が発声する一つの区間と定義する。

2.5 発話群からの会話抽出

発話を音声認識しテキスト化した発話内容のデータを元にし、各発話の中の語彙から話者の意図を抽出する技術は先行研究において明らかにされ、対話システムや対話エージェントなどに活用されている[35-38]。例えば、この技術を利用し、「食事場所を探している」、「明日の天気を知りたい」といったような話者の意図を抽出することができる。

また、2.3で示したようにグループにおける意思決定への支援を実現するために、会話を行っているグループの各メンバー間の関係性を把握してグループ種別を判別する。その種別ごとにみられるグループ特徴から当該グループの状態を推定する。その手順としては、まず、1対1のメンバー同士の発話を抽出して、各メンバー間の関係

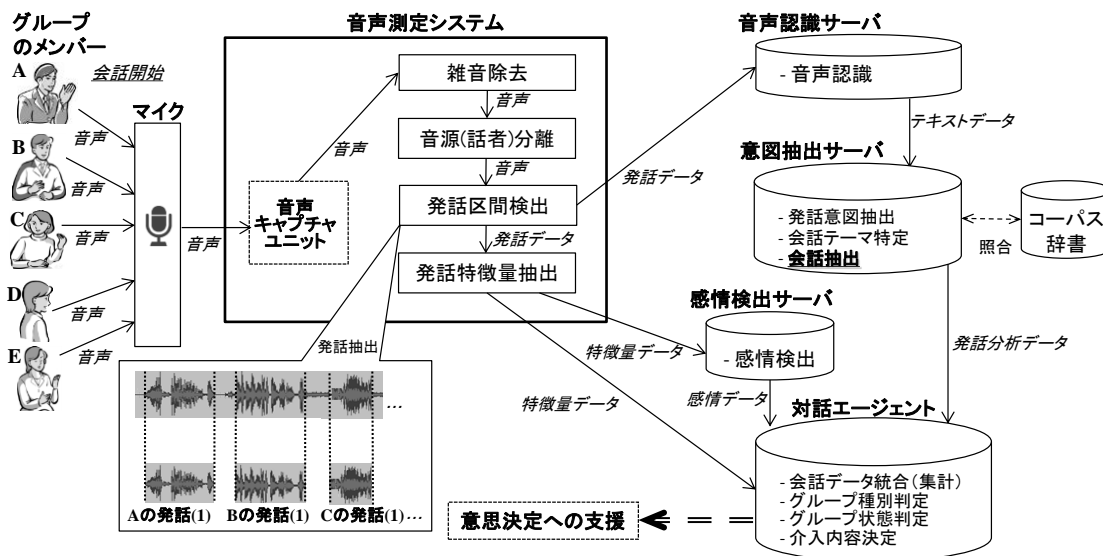


図2 会話抽出関連のシステム全体構成

性を把握し、そして、グループに属する全メンバーの発話群を抽出して、それらからグループ状態を推定する流れを考えている。

このために、参加メンバーや該当する発話を特定して、会話を切り出し、複数の会話が同時に行われる場合でも、同一の会話に含まれる発話を正確に特定することが望まれる。もし、発話内容に着目せず、時間等によって機械的に発話群を切り出すと、異なる会話に含まれる発話も一緒に切り出され、関係のないメンバーや発話意図までが同一テーマの会話に紐づけられてしまう。

会話を切り出す際には、前述の発話意図抽出の技術を活用し、個々の発話での話者の意図を明らかにすると共に、同一テーマと推定される一連の発話を抽出して、一つの会話の開始から終了までに含まれる発話を特定する。(この手法の概要は3章に示す)

そして、この発話意図抽出による手法により明らかにした発話内容と、2.4で述べた発話特徴量とを紐づけて分析し、2.2で示したグループダイナミクスの考え方を適用して、グループ種別や状態を推定し、前述したような意思決定支援につなげる。この手法により、複数の会話が同時に行われる場合でも、推定された各発話のジャンル、話題となっている場所、意図に着目して整理し、いずれの発話が同一の会話に含まれるかを正確に特定することができる。

3. 提案手法

2章で述べた関連研究等をふまえ、グループにおける会話音声モニタし、各話者の発話特徴量測定、音声認識、発話語彙の意図抽出、発話状況集計を行った上で、発話群の中から一つの会話の中に含まれる発話を抽出する手法を以下のように考えた。この手法において、会話抽出に関連するシステム全体の構成を図2に、会話に含まれる発話を抽出するまでの流れを図3に示す。

なお、本提案で対象とするグループは、互いに近接する場所において会話を行う2~10人程度の小規模なグループを主に想定し、1グループが複数のサブグループに分かれて同時に会話することがあるものとして取り扱う。

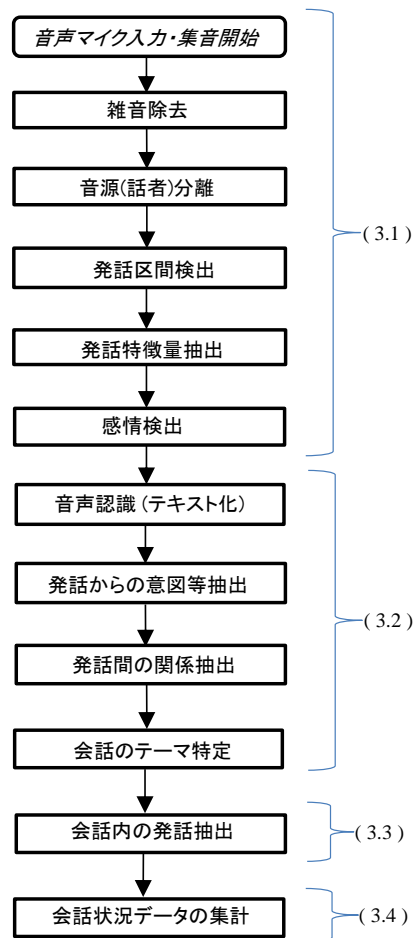


図3 会話内の発話抽出までの流れ

3.1 発話特徴量の抽出

まず、マイクに入力された音声から雑音を除去する。本手法を自動車内で会話するグループへ適用することを目指しているため、発話区間検出や音声認識が適切に実施できるよう、自動車走行時に発生する雑音を除去する必要がある。そして、複数の発話が同じタイミングで重なって発生するような場合でも、個別に発話を分析できるように、話者ごとに音声を分離して抽出する。

次に、話者ごとに分離された音声から発話区間を検出し、検出した発話区間における速度や音圧レベルといった特徴量を抽出する。この発話特徴量の変移を分析して、当該話者の感情を検出し、これらの結果を発話内容と紐づけて分析する。この後段で、会話の進行と共に、発話の特徴量がどのように変化しているのかを把握し、各メンバーの状態、そしてグループ全体の状態を推定する処理につなげる。

3.2 発話意図の抽出

発話区間を検出して抽出された発話音声を音声認識し、各発話の内容をテキスト化し、発話意図の抽出等を以下の通り行う。

- (1) 音声認識によって得られた各発話のテキストの文字列を、「ジャンル」(飲食, 旅行, 音楽, 天候等), 「話題となっている場所」(地名, ランドマーク, 店舗・施設名等), 「意図」(話題の切り出し, 提案, 集約等)の各専用辞書と照合する。

専用辞書としては、意図を特定するための「話題の切り出し, 提案, 質問, 肯定, 否定, 集約」といった場合に使われる語彙, 話題となっている場所を特定するための「地名, 施設・ランドマーク等」に関する語彙や、ジャンルを特定するための「飲食, 旅行, 音楽, 天候等」に関する語彙の辞書を用意している。辞書と照合する抽出語の例は、表 1 に示す。

また、発話を、「発話のジャンル G」, 「話題となっている場所 P」, 「発話の意図 I」と照合した結果は、次のように示される。

$$\text{発話 } n(S_x) = (G_n, P_n, I_n)$$

n : 発話 ID=1~k (発話の発生順に発話 ID を付与),

S_x : 話者 (A, B, C...)

例えば、話者 A の発話 1 を辞書と照合し、「G:飲食」, 「P:鎌倉」, 「I:話題の切り出し」がマッチした場合、発話 1(A) = (“飲食”, “鎌倉”, “話題切り出し”) となる。

- (2) いずれかの発話の終了から次に発生する発話までの間隔が一定時間、例えば 1 分を超えている場合は、当該間隔の前の発話までを一つの発話群とし、更に間隔の後の発話からは別の発話群として分割する。

- (3) (1)で明らかにした各発話の「発話の意図 I」に着目し、同じジャンルの発話を中心に意図的なつながりを推定する。例えば、意図が「提案」の発話の後には「肯定」, 「否定」, 「質問」の発話が、「質問」の発話の後には「回答」の発話が発生しやすいという「発話ターン交代の原則」[39]に基づき、意図的につながっている発話群を推定する。

ジャンルが不明であった発話については、この意図的なつながり推定によってジャンルの推定を行う。

表 1 辞書と照合する抽出語の例

辞書種類	意図等	抽出語 (例)	
意図	話題切り出し	~ようよ	
	提案	~は どおー	
	肯定	いいよ	
	否定	~よりは~	
	質問	~どうなのかなあ	
	回答	~は ね~	
	集約	~するよー	
場所	場所・地名	地名	鎌倉
		地域名	関東
		施設・ランドマーク	スカイツリー
	場所に付加される語	現在いる場所	~に入ったね
		これから行く場所	~は楽しみだね
		過去にいった場所	~はよかったね
ジャンル	飲食	料理	イタリアン
		飲物	コーヒー
		飲食店	レストラン
	天候	天気	晴れ
		寒暖	寒い
		天気予報	明日の天気
	スポーツ	競技	テニス
		スポーツイベント	運動会
	音楽	カラオケ	カラオケ
		歌	歌う

- (4) (2)において時間間隔で分割し、更に(3)で明らかにした各発話のジャンル、意図に基づいた発話のつながりを利用し、同一ジャンルの発話群に分ける。

- (5) 同一ジャンルの発話群の中で、その発話の「意図 I」が「話題切り出し」とされる発話を同一テーマ会話の開始の発話、「意図 I」が「集約」とされる発話を同一テーマ会話の終了の発話と推定する。

但し、会話開始の発話と推定した発話の前に、分類されていない同一ジャンルの発話がある場合はそれらの発話も当該会話と推定し、また会話終了の発話と推定した発話の後に、分類されていない同一ジャンルの発話がある場合はそれらも当該会話と推定する。切り出された一つの会話は、その会話に含まれる発話によって、次のように示される。

$$\text{会話 } m(S_A, S_B, S_C \dots) = \{\text{発話 } 1(S_A), \text{発話 } 2(S_B), \text{発話 } 3(S_C) \dots\} = T_m\{(G_1, P_1, I_1), (G_2, P_2, I_2), (J_3, P_3, I_3) \dots\}$$

m : 会話 ID = 1~k (会話の発生順に会話 ID を付与),
 T_m : 単一の会話テーマ, S_x : 話者 (A, B, C...)

例えば、話者 A, B, C のテーマ「飲食」の発話群が「会話 1」に切り出される場合は、次のように示す。

$$\text{会話 } 1(A, B, C) = T_{\text{“飲食”}}\{(\text{“飲食”, “na”, “話題切り出し”}), (\text{“飲食”, “鎌倉”, “提案”}), (\text{“飲食”, “na”, “提案”}) \dots\}$$

na: not applicable

- (6) (5)で同一テーマの会話における開始/終了の発話が推定できなかった場合は、(4)で同一ジャンルとして切り出された連続する発話群を同一テーマの会話と推定する。

(7) 発話の中にジャンル等を抽出するのに用いるキーワードが含まれていない場合、前後の時間に発生している発話で、会話の開始/終了として推定される発話の間に発生している同じ話者の発話は、同じ会話に含まれると見なして、該当発話をいずれかの会話の発話として分類する。

例えば、「発話 9(A) (“na”, ”鎌倉”, ”na”)」となる場合、その前に発生している同じ話者の「発話 1(A) (“飲食”, ”na”, ”話題切り出し”)」を参照して、「発話 9(A)」を「発話 1(A)」と同じ会話に含まれる発話と分類する。

3.3 発話群からの会話抽出

各発話について意図抽出した結果を集計し、個々の発話のジャンル、話題となっている場所、意図を明らかにすることによって、同一テーマと推定される一連の発話を選別し、会話を抽出する。

例えば、表 2 と図 4 に示すように 5 人の話者によって会話が行われた場合、その発話内容（音声認識したテキストデータ）から抽出される語（キーワード）を専門辞書と照合して得られた意図等を集計し、表 3 と図 5 に示すように、同一テーマと推定される 3 つの会話（発話群）に分けて整理する。

グループの全員が、順次、発話している場合でも、全員が同じ会話に参加しているとは限らないが、そのような場合でも、推定された「各発話のジャンル」、「話題となっている場所」、「意図」に着目して整理すれば、いずれの発話が同一の会話に含まれるかを特定することができる。

3.4 会話状況データの集計

3.3 で示したように会話に含まれる発話特定することにより、会話に含まれる各発話の特徴量データ、音声認識した発話のテキストデータと発話意図等の変移、当該話者の感情を紐づけて、会話内のデータを時系列に整理することができる。図 6 に、会話に含まれる各発話データを分析したイメージを示す。

このように会話抽出し、会話状況データを集計することにより、会話の進行と共に発話の特徴量や話者の感情等がどのように変化しているのかを把握し、各メンバーの状態、グループ全体の状態の推定に活用する。そして、1 章で述べたようにグループ状態に応じたグループの意思決定支援へつなげることができると考えられる。

4. 予備実験

提案手法について、その基本的な有効性を確認するため、2 種類のグループによるテスト会話を利用して、予備実験を実施した。

この予備実験では、音声測定システム（発話区間検出、発話特徴量抽出等の機能を持つ）、音声認識サーバ（発話音声のテキスト化機能を持つ）、意図抽出サーバ（発話意図抽出、会話テーマ特定、会話抽出の機能を持つ）で構成されるテスト用の測定・分析環境（図 2 のシステム全体構成において、発話分析までを行う環境）を用意し、

表 2 収集した発話群の例

時刻	発話 ID	話者	発話内容(テキスト)
12:00:00	1	A	昼食場所をどこにするか、決めようよ。
12:00:06	2	B	北鎌倉のイタリアンはどおー。
12:00:10	3	C	イタリアンよりは、由比ヶ浜海岸沿いの海鮮料理がいいなあ。
12:00:12	4	D	明日の天気はどうかなあ。
12:00:17	5	A	そうだね。海の近くにだから、魚がいいね。
12:00:19	6	E	明日は晴れるって、天気予報がでているよ。
12:00:25	7	B	わかった。じゃあ、海鮮料理でいいよ。
12:00:27	8	D	それはいい。明日の運動会は心配ないね。
12:00:30	9	A	じゃー、由比ヶ浜へ向って出発するよー。
12:00:40	10	A	ところで、今晚、カラオケに行かないかい？
12:00:45	11	B	それいいね、久しぶりに歌ってみたいね。

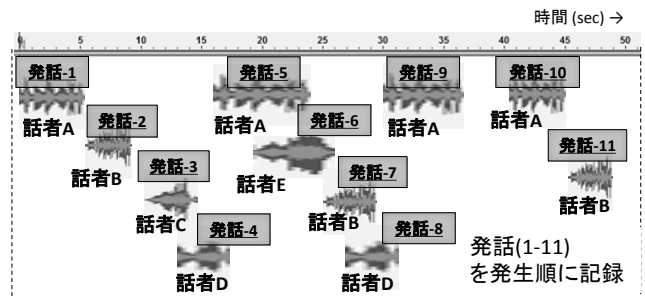


図 4 収集した発話群の発生イメージ

表 3 発話群からの会話抽出例

発話 ID	話者	ジャンル	話題の場所	意図	テーマ等
1	A	飲食(昼食)	—	話題切り出し	(会話-1) テーマ: 飲食 場所: 鎌倉
2	B	飲食(料理)	北鎌倉	提案	
3	C	飲食(料理)	由比ヶ浜	否定/提案	
5	A	飲食(料理)	海	肯定	
7	B	飲食(料理)	—	肯定	
4	D	天候	—	質問	(会話-2) テーマ: 天候 (天候×スポーツ)
6	E	天候	—	—	
8	D	スポーツ(運動会)	—	肯定	(会話-3) テーマ: 音楽
10	A	音楽(カラオケ)	—	話題切り出し, 提案	
11	B	音楽(歌)	—	肯定	

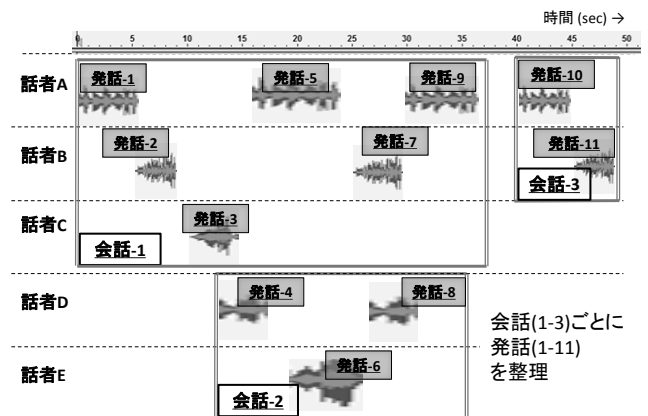


図 5 発話群からの会話抽出イメージ

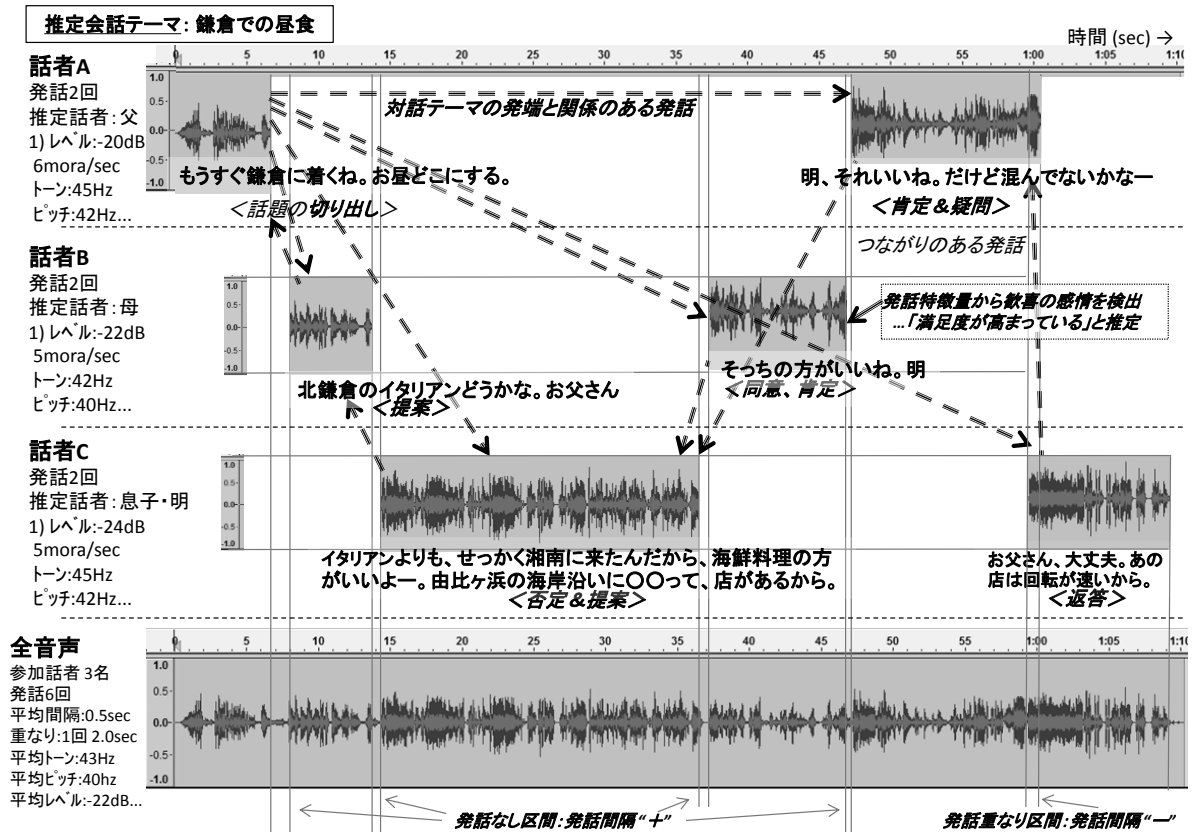


図6 会話内発話の分析イメージ

表4 予備実験の結果

グループ	会話テーマ	発話区間検出 検出数 (検出率)	対象語音声認識 認識数 (認識率)	発話意図等抽出 抽出数 (抽出率)	会話内発話抽出 抽出数 (抽出率)	開始発話 抽出可否	終了発話 抽出可否
1	1	69/74 (93.3%)	19/29 (65.5%)	19/69 (27.5%)	69/69 (100.0%)	OK	OK
	2	65/70 (93.2%)	20/25 (80.0%)	20/65 (30.8%)	65/65 (100.0%)	OK	OK
2	1	45/47 (95.7%)	18/21 (85.7%)	18/45 (40.0%)	45/45 (100.0%)	OK	OK
	2	61/65 (93.8%)	21/25 (84.0%)	21/61 (34.4%)	61/61 (100.0%)	OK	NG

30歳台から50歳台の3人の男性で構成されるグループ(グループ-1)と、20歳台から30歳台の2人の男性で構成されるグループ(グループ-2)によって、テスト会話を同時に行い、その音声を測定・分析した。

なお、提案手法は前述のとおり、自動車等で移動するグループへの適用を検討しているが、現段階では車内雑音を十分に抑制するシステムの実装に至っていないため、本予備実験は雑音条件が緩やかな会議室内の環境で実施した。

テスト会話は、グループ-1がスポーツイベント(テーマ-1)、東京観光(テーマ-2)について、グループ-2が納涼大会(テーマ-3)、宴会の余興(テーマ-4)について、各テーマ約5分間、それぞれ連続して行った。会議室内で実施し、長テーブル席の左側に3人のグループが、右側に別の2人のグループが座った。

マイクは、振動などによる雑音を除去する低周波フィルター内蔵の指向性マイクを、5人の前に配置した。

雑音除去の機能は、マイクの内蔵フィルターを利用し、音源分離は指向性マイクを各話者の近傍に配置することによって実現し、5人の話者ごと各発話音声を収録した。

予備実験を実施し得られた結果は表4に示す通りで、また、実施結果からの考察は次の(1)~(4)の通りとなる。

なお、「発話区間検出」は実際の発話区間を漏れなく検出できた発話数、「音声認識」は意図等の抽出対象語を正確に認識できた発話数、「意図等の抽出」は検出した発話において意図等を抽出できた発話数、「会話内発話抽出」は提案手法によって会話内の発話として抽出できた発話数、「開始/終了発話」は会話における話題の切り出し・集約の発話の抽出可否によって、それぞれの結果を集計した。

- (1) 本予備実験は会議室内の雑音等の影響が比較的に少ない環境で実施したため、入力された音声については、

適切に雑音除去及び音源分離がされ、おおむね発話区間を適切に検出することができた。(発話区間検出率:93~95%)

雑音除去及び音源(話者)分離については、今後想定している雑音状況がより劣悪となる自動車内での会話にも適用できるように、当該機能を向上させることとする。

- (2) 意図等を抽出するための語(キーワード)の音声認識率は65~85%で、音声認識率としては許容できる範囲と考えられる。

しかし、発話区間検出後に音声認識サーバへ音声ファイルを送信して、音声認識(テキスト化)を実施する手順となっていたため、当該発話終了後、音声認識結果を取得するまでに、当該発話時間分の遅延が発生していた。

今後は、音声ストリームを速やかに音声認識サーバへ送信して処理する手順を実現し、リアルタイムでの会話への介入に活用できるように、遅延を改善することとする。

- (3) 意図等を抽出できた発話は、検出した発話の27~40%にとどまった。また、開始を判定する語は4つの会話すべてで抽出できたが、会話の終了を判定する語は1つの会話で抽出できなかった。

これらは、意図等を抽出するためのキーワードが含まれていない発話があるためによるものである。

しかし、時系列な発話の流れを考慮し、意図抽出できた発話を元に意図未抽出の発話を分析することによって、結果的には全ての発話を会話ごとに分離することができた。

今後は、発話意図等の抽出数を増やすために、抽出用の語をより多く、専用辞書に登録し、幅広い種類の会話で、より正確に発話を分離できるように、手法の改善を検討していくこととする。

- (4) 会話に含まれると特定した発話について、発話特徴量と発話意図等を紐づけて、会話内の発話データを集計できることが確認できた。

これにより、例えば、会話の中で、反発が生じた時や、意思決定事項が集約された時のような特定のタイミングの発話特徴量変化を抽出し、発話を利用したグループ状態推定や、その状態に応じたグループにおける意思決定支援への見通しが得られた。今後は、上記(1)~(3)における課題を解決すると共に、意思決定支援の手法を具体化していくこととする。

5. まとめ

我々は、グループダイナミクスや会話分析の先行研究をふまえて、グループの状態に応じて、適切な参考情報やアドバイスを提示するといった介入によって、グループにおける意思決定を支援することを目指しているが、これを実現する上で重要な要素となっている会話抽出について検討し、その方法をまとめた。

提案方法について、テスト会話音声による予備実験を実施した結果、同時に複数の会話が行われる場合においても、発話群の音声の中から、発話を会話単位に分離し、単一の会話に含まれる発話を抽出する実現性が認められた。

但し、発話の音声認識(テキスト化)の結果取得が遅延しているため、処理手順変更によって遅延を解消し、意図等が抽出できている発話が少ないことから、より多く抽出できるように、抽出するためのキーワードの専用辞書への登録を増やす必要があることが明らかになった。

また、同一テーマの会話開始時と終了時の発話を特定することが重要となるので、当該発話を確実に特定する手法等も併せて検討し、判定方法の見直しなどの対策により、全般的な対策を図る。

更に、本稿で示した会話抽出方法を活用して、会話内の各発話の内容と特徴量変化を紐づけて分析することにより、グループの状態推定や、その状態に応じたグループにおける意思決定への支援を想定しているが、今後は、その意思決定を支援する手法を具体化すると共に、フィールドでの検証を予定する。

謝辞

本研究の一部は、株式会社トヨタ IT 開発センターの技術協力を受けて行われた。ここに感謝を表す。

参考文献

- [1] Forsyth, D. R., "Group Dynamics", Thomson Wadsworth, Fourth Edition, (2006).
- [2] Lickel B., Hamilton, D. L., Wierzchowska, G., Lewis, A., Sherman, S. J., Uhles, A. N., "Varieties of groups and the perception of group entitativity", *Journal of personality and Social Psychology*, 78, (2000).
- [3] Walster, E., "The effect of self-esteem on romantic liking", *Journal of Experimental Social Psychology*, 1, (1965).
- [4] Kelly, H. H., Thibaut, J. W., "Interpersonal relations", Wiley, (1978).
- [5] 釘原直樹, "グループ・ダイナミクス: 集団と群衆の心理学", 有斐閣, (2011).
- [6] 集団行動の心理学: "ダイナミックな社会関係のなかで", サイエンス社, (2011).
- [7] Stasser, G., William T., "Pooling of unshared information in group decision making: Biased information sampling during discussion", *Journal of personality and social psychology*, 48, 6, (1985).
- [8] Hogg, M. A., Tindale, S., *Blackwell handbook of social psychology: Group processes*. John Wiley & Sons, (2008).
- [9] Deutch, M., & Gerard, H. B., A study of normative and informational social influence upon individual judgment. *Journal of Abnormal and Social Psychology*, 51, (1955).
- [10] Hinsz, V. B., Thindale, R. S., Vollrath, D. A. The emerging conceptualization of groups as information processors. *Psychological Bulletin*, 121, (1997).
- [11] Sarter, N. B., Schroeder, B., Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43, 4, (2001).

- [12] Herrera-Viedma, E., Martinez, L., Mata, F., & Chiclana, F., A consensus support system model for group decision-making problems with multigranular linguistic preference relations. *Fuzzy Systems, IEEE Transactions*, 13, 5, (2005).
- [13] DeSanctis, G., Gallupe, R. B., A foundation for the study of group decision support systems, *Management science*, 33, 5, (1987).
- [14] 吉原智恵子, 飛田操, "集団討議における集団過程と個体間影響過程の関係", *日本福祉大学情報社会科学論集*, 8, (2005).
- [15] 津村 俊充, "人間関係における「プロセス」を再考する -G. Weinstein, E. H. Schein & W. B. Reddy のプロセスの視点より-", *人間関係研究*, 南山大学, 10, (2011).
- [16] 駒形憲彦, 大塚裕子, "ディスカッションにおける意思決定プロセスの分析(第5回対話システムシンポジウムおよび一般)", *言語・音声理解と対話処理研究会*, 72, (2014).
- [17] Pon-Barry, H. R., Shieber, S. M., "Identifying uncertain words within an utterance via prosodic features", *Proceedings of the 10th Interspeech*, (2009).
- [18] Pon-Barry, H., Shieber, S., "The importance of sub-utterance prosody in predicting level of certainty", *Proceedings of NAACL Human Language Technologies*, (2009).
- [19] Kitaoka, N., Takeuchi, M., Nishimura, R., & Nakagawa, S., "Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems", *人工知能学会論文誌*, 20, (2005).
- [20] 西村良太, 中川聖一, "応答タイミングを考慮した音声対話システムとその評価", *情報処理学会研究報告, SLP, 音声言語情報処理*, 2009, 22, (2009).
- [21] 西村良太, 北岡教英, 中川聖一, "音声対話における韻律変化をもたらす要因分析", *音声研究*, 13, 3, (2009).
- [22] 竹内真士, 北岡教英, 中川聖一, "韻律・言語情報を素性とした決定木による自然な応答生成タイミングの検出", *日本音響学会研究発表会講演論文集*, 1, (2003).
- [23] 加藤弓子, 釜井孝浩, "'Strained Rough Voice'への簡易変換における変調パラメータ", *日本音響学会誌*, 66, 12, (2010).
- [24] 柴崎晃一, 光吉俊二, "抑揚からの感情認識の評価: 感性制御技術(ST)の評価と人間の感情の評価法について", *電子情報通信学会技術研究報告, TL, 思考と言語*, 105, 291, (2005).
- [25] 重永 實, "感情の判別分析からみた感情音声の特性", *電子情報通信学会論文誌, A 83, 6*, (2000).
- [26] Ariga, M., Yano, Y., Doki, S., Okuma, S., "Mental Tension Detection in the Speech based on physiological monitoring", *Systems, Man and Cybernetics, ISIC, IEEE International Conference*, (2007).
- [27] 新庄広, 天野明雄, 石橋厚, 山口宇唯, 内部こなぎ, 桑本英樹, "マルチモーダル対話技術による知的ユーザーインタフェース", *日立評論*, 87, 10, (2005).
- [28] 北岡教英, 赤堀一郎, 中川聖一, "スペクトルサブトラクションと時間方向スムージングを用いた雑音環境下音声認識", *電子情報通信学会論文誌, D 83, 2*, (2000).
- [29] 秋田祐哉, 河原達也, "多数話者モデルを用いた討論音声の教師なし話者インデキシング", *電子情報通信学会論文誌, D 87, 2*, (2004).
- [30] 榊原健一, 石黒浩, 萩田紀博, "Vocal Fry 発声区間の自動検出法", *電子情報通信学会論文誌, D 89, 12*, (2006).
- [31] 後藤真孝, 伊藤克亘, 速水悟, "自然発話中の有声休止箇所のリアルタイム検出システム", *電子情報通信学会論文誌, D 83, 11*, (2000).
- [32] 石塚健太郎, 藤本雅清, 中谷智広, "音声区間検出技術の最近の研究動向", *日本音響学会誌*, 65, 10, (2009).
- [33] 西村雅史, 伊東伸泰, "講義コーパスを用いた自由発話の大語彙連続音声認識", *電子情報通信学会論文誌, D 83, 11*, (2000).
- [34] 中川聖一, "音声認識研究の動向", *電子情報通信学会論文誌, D 83, 2*, (2000).
- [35] 白木将幸, 伊藤敏彦, 甲斐充彦, 中谷広正, "自然発話文における統計的な意図理解手法の検討", *情報処理学会研究報告, SLP, 音声言語情報処理*, 15, (2004).
- [36] Hodjat, B., Amamiya, M., "Applying the Adaptive Agent Oriented Software Architecture to the Parsing of Context Sensitive Grammars", *IEICE transactions on information and systems*, E83-D, 5, (2000).
- [37] Zhong, G., Hodjat, B., Helmy, T., Amamiya, M., "Software Agent Evolution in Adaptive Agent Oriented Software Architecture", *Proceedings of International Workshop on the Principles of Software Evolution (IWPSE99)*, (1999).
- [38] Hodjat, B., Amamiya, M., "Introducing the Adaptive Agent Oriented Software Architecture and Its Application in Natural Language User Interfaces", *Agent-Oriented Software Engineering Lecture Notes in Computer Science*, 1957, (2001).
- [39] Sacks, H., Schegloff, E. A., Jefferson, G., "A simplest systematics for the organization of turn-taking for conversation", *language*, (1974).