音声情報を用いた複数人自由対話における重要文抽出

河原 真太郎[1,a)]　山村 崇[2,b)]　嶋田 和孝[3,c)]

概要：本論文は，複数人自由対話を対象とした重要文抽出において，音声情報を考慮した手法を提案する．自由対話において，重要度の高い発話は声量が大きくなる傾向や，重要文に多く見られる質問文において語尾の音程が高くなる傾向がある．これらのことから，重要文抽出において，音声情報を追加することは精度向上につながると考えられる．先行研究に本手法の素性を追加し，機械学習によって重要文抽出を行うことで，自由対話における音声情報の有効性を検証する．

キーワード：複数人自由対話，重要文抽出，非言語情報，声量，音程

Important utterance extraction using speech information for multi-party conversation

Shintaro Kawahara[1,a)]　Takashi Yamamura[2,b)]　Kazutaka Shimada[3,c)]

Abstract: In this paper, we propose a method of important utterance extraction for multi-party conversations. In spontaneous conversations, speech information has an important role. For example, important utterances are often louder than non-important utterances because speakers want to emphasize the utterances. Besides, a terminal syllable with high voice pitch sometimes indicates the importance of an utterance because questions in conversations are often important utterances. We add new features based on speech information to our previous method. The experimental result shows a marginally significant positive effect on a selected dataset.

Keywords: Multi-party conversation, important sentence extraction, non-verbal, voice, musical interval

1. はじめに

第 3 者が新聞や会議といった膨大な量のテキストの内容を理解したい場合，全文に目を通さなければならず，多くの時間と労力が必要となり，大きな負担がかかる．この負担を軽減させるため，テキストの自動要約の研究が近年行われている．これまでの自動要約に関する研究の多くは，新聞記事や論文のようなテキストを要約対象にし，書き手が 1 人で，かつ文体が書き言葉であるものが多かった [1]．また，要約の手法については，原文から単語の出現頻度や手がかり語，文の出現位置などの情報から重要と判断される文 (以下，重要文) を抽出し，それらを要約とする抜粋要約が主流である [2][3]．本研究でも，この抜粋要約 (重要文抽出) を研究の対象とする．

自動要約に関する研究が進むにつれて，新聞記事や論文のようなテキストだけでなく，複数人の発話で構成された対話文を要約対象とした研究も行なわれるようになってきた．複数人の対話には，一つの題目に関して話者が自由に発言する自由対話がある．自由に発言するというのは，会議やコールセンターといった，決められた枠組みに沿った対話でなく，雑談のような対話を指す．自由対話では，笑いの発生や，盛り上がりの発生，発話のオーバーラップと

[1] 九州工業大学 情報工学部 知能情報工学科
Kyushu Institute of Technology
[2] 九州工業大学 大学院情報工学府 先端情報工学専攻
Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology
[3] 九州工業大学 大学院情報工学研究院 知能情報工学研究系
Department of Artificial Intelligence, Kyushu Institute of Technology
[a)] s_kawahara@pluto.ai.kyutech.ac.jp
[b)] t_yamamura@pluto.ai.kyutech.ac.jp
[c)] shimada@pluto.ai.kyutech.ac.jp

いった現象が起きる．たとえば，盛り上がり状態では，話者が主観的な発話や，相手の発話を補う発話を発しており，話者が対話に興味を示していると考えられる [4]．

本研究室では，自由対話要約の研究において，言語情報と非言語情報の両方を考慮した重要文抽出を行っている．自由対話における各発話の重要性や発話同士の関係性といった言語情報に，笑いの距離，発話速度といった非言語情報を考慮することで，重要文抽出の精度が改善することが示されている [5][6][7]．本論文では，先行研究では触れられていない声量や音程といった音声情報を新たに追加する．発話の特徴として，重要な意見を伝えるときに声量が大きくなることや話題を提供する際の質問文では語尾が上がるといったものが考えられる．このような音声情報を用いた重要文抽出を行うことで，その有効性を検証する．

2. 関連研究

本研究室では自由対話の要約に関する研究をしている．嶋田ら [5] は，複数人の自由対話を構成する各発話の重要性や発話同士の関係性といった言語情報を考慮し，機械学習を用いた重要文抽出を提案している．徳永ら [6] は，嶋田らの提案する言語情報に，笑いの有無や盛り上がりの有無といった非言語情報を追加することで重要文抽出の精度がわずかながら改善することを述べている．徳永らの結果を受けて，山村ら [7] は，発話速度，発話のオーバーラップといった発話時間に関する非言語情報の素性を考慮した重要文抽出を行った．徳永らの考慮した非言語情報の素性と発話時間に関する素性を組み合わせることで，重要文抽出の精度がさらに改善することがわかった．これらのことから，新たな非言語情報を考慮することは，重要文抽出に有効であり，非言語情報の素性を組み合わせることで更なる改善が考えられると述べている．

本研究室での先行研究が触れていない他の非言語情報として，発話の声量，音程といった音声情報がある．発話の特徴として，相手に重要な意見を伝えるときは，声量が大きくなることや，語尾の強調といったことがある．Hsueh ら [8] は，発話の声量，音程，発話速度といった音声情報をベースに，議論の決定事項に関する文を抽出している．山田ら [9] は，非言語情報を用いた発話の重要度の判定において，声量の大きさが重要であることを述べている．さらに，対話において新たな話題を提供する際，質問形式で行われることが多いことから重要度が高くなることが考えられる．この質問文は，語尾の音程が上がることや，重要な発話では音程が低くなるといった特徴がある．また，小林ら [10] は，韻律情報を用いた講演の要約において，要約の精度の向上に貢献することを報告している．本研究ではこれらの研究にならい，自由対話要約への音声情報の有効性を検証する．

3. ベースライン

まず，関連研究で述べた嶋田ら，徳永ら，山村らの手法について説明する．嶋田らは，対話文を構成する各発話の重要性や発話同士の関連性を考慮し，重要文抽出を行っている．手法としては，機械学習による分類器を用いて重要文を抽出している．さらに，徳永らおよび山村らは，嶋田らの言語情報の素性に，笑いの距離や，発話速度といった非言語情報の素性を新たに追加し機械学習を用いて，重要文抽出を行っている．この重要文抽出の結果から，最も精度の良かった組み合わせである 21 個の素性を説明する．これらの素性は，嶋田らの提案する言語情報に関する 4 つの素性群 (ABCD) と，徳永ら，山村らの提案する非言語情報に関する 2 つの素性 (E_1, E_2) に分類される．これらの素性に基づく手法を本研究でのベースラインとする．

素性群 A: 発話単体の特徴に関する素性

発話単体の特徴に関する素性は，ある 1 発話のみの特徴に着目した素性である．発話単体に関する 4 つの素性について説明する．

A_1: 発話の長さ

文長の長い発話は，多くの情報を持っていることが多く，重要文である可能性が高いと考えられる．そこで，発話を構成する形態素数を素性として用いる．

A_2: 複数回出現の有無

発話に含む単語の全発話における出現回数を数える．対話全体において複数回出現する単語を含んでいる発話は重要文である可能性が高い．対話全体で 2 回以上出現している名詞を複数回出現単語とし，発話における複数回出現単語の有無を素性として用いる．

A_3: 用言の有無

自由対話には，用言を含まないような短絡的な発話が多く存在することから，動詞，形容詞の有無を素性として用いることは有効であると考えられる．

A_4: 疑問文の有無

疑問文は対話における要点となりうる可能性が高く，重要性を持っていると考えられる．本手法では，クエスチョンマークを含む発話を疑問文とする．クエスチョンマークは，書き起こしの際に，アノテータの主観によって決められている．

素性群 B: 発話間の特徴に関する素性

発話間の特徴に関する素性は，ある発話とその前後の発話間における特徴に着目した素性である．前後の発話との関連性が大きい発話は重要である可能性が高い．発話間の特徴に着目した 6 つの素性についてそれぞれ説明する．

B_1: 発話の文長差

重要度や関心の高い発話の直後は，その発話への反応

を表すような文長の短い発話がくることが多い．このことから，後方の発話に比べて文長が大幅に長い発話は，高い重要性を持っている可能性がある．そこで，発話自身の形態素数と後方の 3 発話の形態素数の差の合計を素性として用いる．

B_2: 複数回出現単語の有無 (直前の発話)

複数回出現単語を含む発話は重要文である可能性が高いと説明したが，これはその直後の発話の重要度にも影響すると考えられる．そこで，直前の発話における複数回出現単語の有無について着目する．

B_3: 疑問文 (直前の発話)

重要性の高い疑問文がある場合，その直後の発話も疑問に対する回答として重要性が高いと考えられる．

B_4: 同一単語の有無 (直前の発話)

前後の発話に同一単語が含まれている場合，発話同士に関連性がある可能性が高い．そこで，直前の発話に同一の名詞，動詞，形容詞が含まれているかをそれぞれ素性として用いる．

B_5: 同一単語の有無 (後方の 3 発話)

同様に，後方の 3 発話に同一の名詞，動詞，形容詞が含まれているかをそれぞれ素性として用いる．

B_6: 発話者の連続性

発話者が連続しているとき，後方の発話は前方の発話の補足として存在している場合がある．このことから，発話者の連続性は重要度に影響する可能性がある．そこで，発話者の連続性に着目し，直前の発話と発話者が同じであるかについて素性を用いる．

素性群 C：　照応性に関する素性

西川ら [11] は，対話文においては照応関係をもつ発話が多く存在しており，照応する側の発話が要約に含まれる場合は，照応される側の発話も要約に含まれなければならないと述べている．また，多くの発話に照応されている発話は高い重要性を持っており，重要文として抽出する必要があると考えられる．そこで，照応性をもつ表現として，指示表現，接続表現，反応表現の 3 つに着目した素性を用いることにする．

C_{1a}: 指示表現の有無

C_{1b}: 指示表現の有無（後方の 3 発話）

「これ」「そっち」「あそこ」のような指示表現が，発話自身及び後方の 3 発話に出現しているかについて調べる．

C_{2a}: 接続表現の有無

C_{2b}: 接続表現の有無 (後方の 3 発話)

「でも」「しかも」「ただし」のような接続表現が，発話自身及び後方の 3 発話に出現しているかについて調べる．

C_{3a}: 反応表現の有無

C_{3b}: 反応表現の有無 (後方の 3 発話)

「あー」「へえ」「うん」のような反応表現が，発話自身及び後方の 3 発話に出現しているかについて調べる．

素性群 D：　キーワード評価値に関する素性

展望台システム [12] を参考に 3 つのキーワード評価値を機械学習の素性として用いる．展望台システムについて簡単に説明する．展望台システムでは，単語の出現頻度によるキーワードを手がかりとして，文章の主題となるキーワードを単語間の関連に基づいて抽出し，低頻度のキーワード中からも文章の主題に合致し，かつ，文章を特徴付けるキーワードを取り出す．展望台システムにおける 3 つのキーワード評価値をそれぞれ説明する．

D_1: 周辺キーワードの評価値

名詞，動詞，形容詞の出現回数を単語ごとにカウントする．そして，対話中に 2 回以上出現する名詞，及び 1 回以上出現する動詞，形容詞を周辺キーワードと定義し，出現回数を与える．この評価値を式 (1) とする．

$$key1(w) = frequency(w) \tag{1}$$

式 (1) を用いて，各文 T に対し，周辺キーワードの評価値として式 (2) を与える．ただし，単語 w が動詞または形容詞であった場合は，補正として「周辺キーワードである名詞の対話中での総出現回数 / 動詞 (形容詞) の対話中での総出現回数」を乗算する．

$$sentence1(T) = \sum_{w \subset T} key1(w) \tag{2}$$

D_2: 中心キーワードの評価値

周辺キーワードが現れた時に，同時に現れやすい単語として中心キーワードを定義する．対話中に出現する各名詞に，式 (3) の評価関数で与えられる評価値を与え，その評価値が上位 10 %以内である名詞を中心キーワードとする．周辺キーワード集合を G，周辺キーワード g を含む発話の数を $n(g)$，名詞 w と周辺キーワード g が同時に出現する発話の数を $n(w \cap g)$ とするとき，名詞 w の評価値は次式となる．

$$key2(w) = \sum_{g \subset G} \frac{n(w \cap g)}{n(g)} \tag{3}$$

式 (3) を用いて，各文 T に対し，中心キーワードの評価値として式 (4) を与える．

$$sentence2(T) = \sum_{w \subset T} key2(w) \tag{4}$$

D_3: 特徴キーワードの評価値

対話の主題となりうる中心キーワードの流れに沿い，かつ文章を特徴付けているような単語として特徴キーワードを定義する．対話の中で重要な役割を果たすキーワードは，何度も出てくる単語だけではない．一度しか出てこない単語の中にも，重要なキーワードは

存在していると考えられる．特徴キーワードの評価値は，中心キーワードが出現するときのみ同時に出現するような単語に着目しており，どれだけ多くの中心キーワードと密接に関係しているかを表す．対話中に出現するすべての名詞に，式 (5) の評価関数で与えられる評価値を特徴キーワードの評価値として与える．中心キーワード集合を S，名詞 w を含む発話の数を $n(w)$，名詞 w と中心キーワード s が同時に出現する発話の数を $n(w \cap s)$ とするとき，名詞 w の評価値は次式となる．

$$key3(w) = \sum_{s \subset S} \frac{n(w \cap s)}{n(w)} \tag{5}$$

式 (5) を用いて，各文 T に対し，特徴キーワードの評価値として式 (6) を与える．

$$sentence3(T) = \sum_{w \subset T} key3(w) \tag{6}$$

これらの式 (2)(4)(6) は，すべて評価値を 1 から-1 の範囲で正規化して利用する．

素性群 E:　非言語情報に関する素性

自由対話の特徴として，コールセンターや会議などではあまり見られない，笑いの発生，発話時間情報などの非言語情報が挙げられる．このような，自由対話特有の特徴である非言語情報を考慮することは，自由対話における重要文抽出において重要であると考えられる．そのなかでも，最も精度が良くなった組み合わせである笑いの距離，発話速度の素性を説明する．

E_1:　external な笑いからの距離

笑いの存在する場面では，話者は意味のある発話をしていると仮定している．嶋田ら [13] は，笑いを外部からの発話 (行動) に対して発生する笑い (external) と自発的な笑い (internal) に分類し，external な笑いの周辺には，笑いを引き起こした原因となる発話が存在するとしている．起こった笑いが external の場合，周辺には笑いを引き起こした発話があるため，その直前 3 発話以内の発話に対し，external な笑いを含む発話からの距離を素性として割り振る．

E_2:　発話速度

発話速度は発話の速さを表す．藤原ら [14] は，対話意図を考慮した対話リズムの分析を行なっており，対話の応答において重要度の高い発話は，発話速度がゆっくりになる傾向があると述べている．このことから，発話速度が遅い発話，ゆっくり強調された発話には，重要な発話が多いと考えられる．式 (7) で得られる発話速度の値を求め，相対的に発話速度が早いか遅いかの 2 値を素性として用いる．また，発話文には漢字で表記されているため，漢字を平仮名に変換し文字数を計算する．

$$\text{発話速度} = \frac{\text{発話文の文字数 (句読点を除く)}}{\text{発話時間}} \tag{7}$$

最後に，以上の 21 個の素性を機械学習に適用し重要文の抽出を行なう．機械学習のアルゴリズムには，多くの分類問題において優れた汎化性能を持つ Support Vector Machine [15](以下，SVM) を用いる．

4. 提案手法

先行研究 (ベースライン) の重要文抽出において，山村らは，新たな非言語情報を追加することで精度が改善すると述べている．そこで，先行研究では触れられていない新たな非言語情報として，発話の声量や音程といった音声情報の素性を提案する．自由対話において，場が盛り上がっているときや，議論が白熱している時，話者は自然と声量が大きくなることが考えられる．さらに，質問をするとき，話者は語尾が上がるといった特徴が見られる．これらのように，疑問文や議論が白熱して声量が大きくなるような状況で，重要文が多く含まれる傾向があることから，音声情報は有効に機能すると考えられる [*1]．

そこで本論文では，音声や音程を用いた素性を新たな非言語情報の素性群 G として先行研究に追加する．この素性について，以下で詳しく説明する．

素性群 G:　音声情報に関する素性

音声情報を用いるにあたって，各発話の声量，音程の値を取得する．声量値は，各発話の発話時間内の波形から振幅の大きさを Root Mean Square（以下，RMS）を用いて求める．RMS とは，実効値のことを指し，RMS を用いることで，発話の声量値を聴感上の声量に一致させることができる．式 (8) を用いて，振幅の実効値を求め，実効値を 声量値$_{ALL}$ とする．

$$\text{声量値}_{ALL} = RMS[x] = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)} \tag{8}$$

ここでは，各発話の振幅の値を x とし，振幅値の数を N とする．

次に，振幅の数によって発話の声の音程を求める．各発話の振幅数を数え，各発話の発話時間を用いて発話の単位時間当たりの振幅数を式 (9) で求め，音程値$_{ALL}$ とする．

$$\text{音程値}_{ALL} = \frac{\text{振幅数}}{\text{発話時間}} \tag{9}$$

さらに，語尾の強調や語尾の音程の素性を提案するにあたり，語尾の声量値，音程値を求める必要がある．そこで，図 1 のように各発話の発話時間を 5 分割にし，

[*1] 3 節のベースラインの説明を見ればわかるように素性群 A や B に疑問文の有無の素性がある．嶋田らは主観的に疑問文を決定しているが，提案手法では音声情報を用いることで客観的に疑問文を決定する点が異なる．

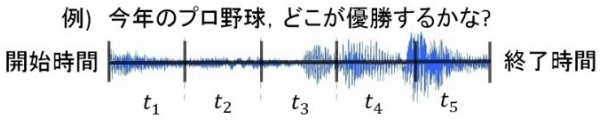

図 1 発話時間を 5 分割した時.

発話の開始時間からそれぞれ分割毎に t_1〜t_5 とする．例えば，図 1 では，t_1 の時間内では，発話は“今年の”の位置になり，t_5 の時間内では，発話は“かな?”の語尾の位置に相当する．分割した t_1〜t_5 の発話の発話時間内での 声量値$_{t1}$〜声量値$_{t5}$，音程値$_{t1}$〜音程値$_{t5}$ をそれぞれ求める．分割した声量値は，分割した時間内の振幅値から RMS を用いて求め，音程値は，単位時間あたりの振幅数を分割した時間内から求める．これらの声量，音程の値を用いて，音声情報に関する素性を以下に提案する．

G_1：**各発話の声量の大きさ**

山田ら [9] は，重要な話の場面では発話の声量が大きくなると述べている．そこで，各発話の 声量値$_{ALL}$ が 1 対話中の全発話の 平均声量値$_{ALL}$ より大きい発話を 1，平均以下の発話を 0 とした素性として用いる．

G_2：**各発話の語尾の声量の強調**

話者は，相手を説得するとき，自信を持って話をするときに語尾を強調する傾向がみられる．自信をもって言い切っている発話は重要文に含まれやすいと考えられる．各発話の 声量値$_{t5}$ が 1 対話中の全発話の声量値$_{t5}$ の平均声量値より大きい発話を 1，平均以下の発話を 0 とした素性として用いる．

G_{3a}：**各発話の語尾の音程上がり (全体平均比較)**

G_{3b}：**各発話の語尾の音程上がり (発話内比較)**

重要文に含まれやすい質問文は，語尾の音程が高くなる傾向が見られる．また，小林ら [10] は，講演の要約において，音程の情報を考慮することで，要約の精度が向上したと述べている．このことから，複数人自由対話においても音程は，有効であると考えられる．

音程値$_{t5}$ の値が，1 対話中の全発話の 音程値$_{t5}$ の平均より高い発話を 1，平均以下の発話を 0 とした素性を G_{3a} とする．また，G_{3a} は対話内の全発話から語尾の音程の高さを比較しているが，発話内のみでも語尾の音程の高さを比較する．音程値$_{t4}$ と比較し，音程値$_{t5}$ が 音程値$_{t4}$ の値より高い発話を 1，音程が低い発話を 0 とした素性を G_{3b} とする．

G_4：**語尾の音程下がり**

質問文に対する応答や同意といった発話は極端に語尾の音程が低くなる傾向がみられる．そこで，各発話の 音程値$_{t5}$ の全体割合を式 (10) で求める．

$$\text{音程値}_{t5}\text{の全体割合 (\%)} = \frac{\text{音程値}_{t5}}{\text{音程値}_{ALL}} \qquad (10)$$

1: D SNS、なんかやってます？
2: B やってる。
3: A facebook とか？
4: B facebook、twitter、mixi。
5: C あー、やってますね。
6: A やりまくりやな。
7: D 俺も、ツイッ、twitter ですね。
8: D アメーバピグは、SNS に入れていいんすかね？
9: B ま、一応いいんじゃない。
10: C ま、入れていいんじゃない。
11: A アメーバピグって最近なんかなかった？事件みたいな。
12: B ないんじゃない特に。あったっけ？あれ？事件？
13: C は、ま、ないかなー。
14: A あのほら、小学生とか中学生が、なんかこう。
15: A 何だろう。十一から三十五歳の、男女の、こうアカウントを盗んで。

図 2 “SNS について”の対話内の一例.

1: A じゃあ、えっと、まあ、今年のプロ野球で。
2: A なんだろう、セリーグとパリーグが最近、何か。
3: A 何か一位が決まったとかそういう。
4: A 何があったのかな？
5: B ああ、そうですね。
6: C もうパリーグは、ホークスが、大分早くから。
7: A 決まってて。
8: D あの、中日はどうなったんですか？
9: C 中日が結局優勝した。
10: B 中日が今日優勝した、うん。
11: A 優勝した？
12: B ようやく、昨日優勝した。
13: D 逆転勝ちみたいでしたね。
14: B 逆転、十…。
15: A はー、ヤクルトは？

図 3 “今年のプロ野球について”の対話内の一例.

発話時間を 5 分割にしていることから，音程値$_{t5}$ の全体割合が 20 %のとき平均音程値の基準になり，音程が低くなる場合は音程値の全体割合が 20 %を下回る．この素性では語尾の音程の低い発話を対象とするため，今回は閾値として 17 %という値を人手で設けた．閾値 17 %を下回った発話を 1，閾値以上の発話を 0 とする．

5. 実験と考察

4 節で説明した新たな素性を先行研究である山村らの手法に追加し，機械学習による複数人自由対話における重要文抽出実験を行なった．実験対象データは，先行研究と同じく 4 名の話者が 1 つの話題について自由に話している 8 つの対話 (全 1295 発言) とした．話題は，“SNS について”，“好きなゲームハードについて”，“好きなファストフード店について”，“今年のプロ野球について”，好きな映画について”の 5 つの話題である．図 2 および図 3 に対話の発話例を示す．A〜D のアルファベットは話者を指し，下線が引かれている発話は重要文 (正例文) となる．

1 対話中の正例文数を ap，抽出した結果中の正例文数 p，

表 1 重要文抽出実験結果 (8 対話).

組み合わせパターン		F 値
(1)	ベースライン (素性群 ABCDE)	0.7976
(2)	(1) + G_1	**0.7991**
(3)	(1) + G_2	0.7942
(4)	(1) + G_{3a}	0.7967
(5)	(1) + G_{3b}	0.7973
(6)	(1) + G_4	0.7941
(7)	(1) + 素性群 G	0.7833

表 2 各対話における重要文抽出結果 (8 対話).

		対話 1	対話 2	対話 3	対話 4	対話 5	対話 6	対話 7	対話 8
(1)	ベースライン	0.833	0.789	0.832	0.835	0.725	0.840	0.749	0.778
(2)	(1) + G_1	**0.833**	**0.796**	0.810	**0.835**	**0.731**	**0.840**	**0.756**	**0.792**

表 3 各対話における発話の声量値の標準偏差.

対話 1	対話 2	対話 3	対話 4	対話 5	対話 6	対話 7	対話 8
251.04	155.81	50.48	86.57	253.16	70.20	107.08	72.05

表 4 重要文抽出実験結果 (7 対話).

組み合わせパターン		F 値
(8)	ベースライン (素性群 ABCDE)	0.7878
(9)	(8) + G_1	**0.7908**
(10)	(8) + G_2	0.7872
(11)	(8) + G_{3a}	**0.7912**
(12)	(8) + G_{3b}	**0.7887**
(13)	(8) + G_4	0.7860
(14)	(8) + 素性群 G	**0.7944**†

抽出した結果数を S とすると，適合率と再現率は式 (11) のように表され，F 値は式 (12) のように表される．この評価尺度では，それぞれの分類器が抽出した重要文に，どの程度正しく正例文が含まれているかが評価される．実装には，データマイニングツール Weka[*2] の SVM[15] を用いた．

$$\text{適合率} = \frac{p}{S},\quad \text{再現率} = \frac{p}{ap} \tag{11}$$

$$\text{F 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \tag{12}$$

8 対話中 1 対話をテストデータとする 8 対話交差検定を用いて，得られた 8 回の結果を平均して F 値を算出した．結果を表 1 に示す．先行研究であるベースラインの素性群の重要文抽出の結果を (1) ベースラインとする．(1) ベースラインより精度が上がった結果は太字で示す．

表 1 の (1) と比較すると，(2) で示す各発話の声量値の大きさの素性を追加した時のみ，F 値を上回ったことがわかる．そこで，(1) のベースラインと (2) の各発話の声量の大きさに関する素性について t 検定を行ったが，有意水準 5 %での有意差は見られなかった (t=2.36，df=7，0.05<p)．

ここで，精度が上がらなかった要因を検証するため，(1) のベースラインと提案手法で精度が一番高くなった (2) の実験結果について対話ごとの結果を比較する．表 2 に (1) のベースラインと (2) 各発話の声量値の大きさの素性を組み込んだ手法に対する 8 対話それぞれの重要文抽出の F 値を示す．表 2 より，声量に関する素性は，対話 3 のみが先行研究より F 値が下がっていることがわかる．対話 3 の重要文抽出の精度が下がった要因を検証するため音声データを確認した．すると，対話 3 では，他の対話に比べ，各発話の声量値が対話内の平均声量値に近く，声量の変化が少ないことがわかった．

そこで，発話の声量値が平均声量値に近い発話がどの程度存在するかを具体的に他の対話と比較するため，対話中における発話の声量値の標準偏差を求めた．その結果を表 3 に示す．表 3 より，対話 3 が他の対話に比べ，標準偏差が小さいことから，対話 3 では発話の声量値が平均声量値に近く，同等の声量値の発話が多くなっていることがわかる．このことから，対話 3 では，音声の特徴量が正しく抽出されず，その結果重要文抽出の精度を下げたと考えられる．

次に音声情報に対する正確な有効性を検証するために，この対話 3 を除いた 7 対話での重要文抽出実験を行った．その結果を表 4 に示す．先行研究の素性群の重要抽出結果は (8) ベースラインに示す．表 4 より，(8) のベースラインと比較すると，(9) の各発話の声量の大きさに関する素

[*2] http://www.cs.waikato.ac.nz/∼ml/weka/

表 5　各対話における重要文抽出結果 (7 対話).

	対話 1	対話 2	対話 4	対話 5	対話 6	対話 7	対話 8
(8)　ベースライン	0.838	0.783	0.829	0.719	0.820	0.762	0.764
(9)　(8) + G_1	0.821	**0.789**	**0.840**	**0.720**	**0.826**	**0.762**	**0.778**
(11)　(8) + G_{3a}	**0.839**	0.763	**0.841**	**0.735**	0.808	**0.782**	**0.771**
(12)　(8) + G_{3b}	0.815	**0.789**	**0.829**	**0.735**	**0.827**	**0.762**	**0.764**
(14)　(8) + 素性群 G	**0.845**	0.776	**0.830**	**0.735**	**0.821**	**0.770**	**0.778**

性，(11)，(12) の語尾の音程上がりに関する素性，(14) の音声情報に関する素性群を追加した場合 (8) より上回ったことがわかる．これらの違いを検証するため，表 5 に表 4 の (8) のベースライン，(9) の各発話の声値の大きさに関する素性，(11)，(12) の語尾の音程上がりに関する素性，(14) の素性群 G をそれぞれ追加したときの，対話ごとの重要文抽出結果を示す．表 5 より，ほとんどの対話において，(8) よりも F 値が上回っていたが，逆に F 値が下回っている対話も存在していた．そこで (8) に対して有意差があるかそれぞれ両側検定の t 検定を行った．表 5 の (8) と一番精度の良かった (14) について t 検定を行ったところ，有意水準 5 %で有意傾向であることがわかった (t=2.44, df=6，$0.05< p <0.1$)[*3].

これらの実験結果から，対話 3 のような声量値に変化があまり生じないデータでは，提案手法は適切な学習が困難であり，有効に機能しないことがわかった．一方で，声量値にある程度の変化がある場合は，表 5 より一定の有効性がみられている．音声情報が機能するか機能しないかは対話全体での声量値に変化の程度に基づいており，これは表 3 の標準偏差などの値で機械的に判断が可能であると考えられる．今後はこのような何らかの基準によって，音声情報に関する素性を適切に使い分け，全体的な精度を向上させる手法の確立が必要である．

6. おわりに

本論文では，自由対話における重要文の抽出実験を行い，声量に関する素性と，語尾の音程に関する素性を新たな非言語情報として先行研究の素性に追加した．対話内で各発話の声量値にある程度変化がみられるといった条件下であれば，一定の有効性が確認された．

今後の課題としては，声量値，音程値をより正確に取得することが考えられる．提案手法では，発話の語尾の声量値や音程値を，発話時間を 5 分割するといった大きい括りで取得していたが，発話の長さによって分割数を変えることでより正確に求めることができると考えられる．また，ピッチの勾配を用いることで音程値を正確に求められると考えられる．他にも，今回触れられていない新たな非言語情報を取り入れることや既存の素性と上手く組み合わせることが精度改善に重要と考えられる．

[*3] 表 5 の (9)(10)(11) については有意差なし．

謝辞　本研究は科研費の助成 26730176 を受けたものです．

参考文献

[1] 奥村学, 難波英嗣. 知の科学　テキスト自動要約. オーム社, 2005.
[2] Tsutomu Hirao, Hideki Isozaki, Eisaku Maeda and Yuji Matsumoto. Extracting Important Sentences with Support Vector Machines. Proceedings of COLING 2002, pp. 342-248, 2002.
[3] Miels Osborne. Using Maximum Entropy for Sentence Extraction. Proceedings of the ACL-02 Workshop on Automatic Summarization, pp. 1-8.
[4] 徳久良子, 寺嶌立太. 雑談における発話のやりとりと盛り上がりの関連. 人工知能学会論文誌, Vol. 21, No. 2, pp. 132-142, 2006.
[5] Kazutaka Shimada, Shinpei Toyodome, and Tsutomu Endo. Conversation summarization using machine learning and scor-ing method. Proceedings of PACLING 2013, 2013.
[6] Yo Tokunaga and Kazutaka Shimada. Multi-party conversation summarization based on sentence selection using verbal and nonverbal information. Proceedings of ICSCAI 2014, 2014.
[7] 山村崇, 徳永陽, 嶋田和孝. 時間情報とテキストセグメンテーションに基づく複数人対話要約手法. 電子情報通信学会, 言語理解とコミュニケーション研究会 (NLC), NLC2015-8, pp. 43-48, 2015.
[8] Pei-Yun Hsueh and Johanna Moore. What decisions have you made: Automatic decision detection in conversational speech. Proceedings of In NAACL/HLT, 2007.
[9] 山田博文, 松田和彦, 田口亮, 桂田浩一, 小林聡, 新田恒雄. 講義再現システムにおけるスライド重要度抽出. 人工知能学会論文誌, Vol. 17, No. 4, pp. 481-489, 2002.
[10] 小林聡, 山口優, 中川聖一. 表層的言語情報と韻律情報を用いた講演音声の重要文抽出. 自然言語処理, Vol. 12, No. 5, pp. 43-68, 2005.
[11] 西川仁, 長谷川隆明, 松尾義博, 菊井玄一郎. 文外照応を含む文の検出による抽出型要約の品質向上. 言語処理学会第 17 回論文集, pp. 216-219, 2011.
[12] Wataru Sunayama and Masahiko Yachida. A Panoramic View System for Extracting Key Sentences with Discovering Keywords Featuring a Document. Systems and Computers in Japan, Vol. 34, No. 11, pp. 81-90, 2003.
[13] 嶋田和孝, 楠本章裕, 横山貴彦, 遠藤勉. 複数人談話における笑いの情報を考慮した盛り上がり判定. 電子情報通信学会, 言語理解とコミュニケーション研究会 (NLC), NLC2012-7, pp. 25-30, 2012.
[14] 藤原敬記, 伊藤敏彦, 荒木健治. タスク指向対話における相互の対話意図を考慮した対話リズムの分析, 人工知能学会言語・音声理解と対話処理研究会. SIG-SLUD-A701, pp. 45-50, 2007.
[15] Vladimir Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag, 1995.