J-012

笑い声を利用した盛り上がりのリアルタイム検出

Detection of Active Parts of Conversation in Real Time Using Laughter

今吉 晃 †　　棟方 渚 †　　小野 哲雄 †

Akira Imayoshi　Nagisa Munekata　Testuo Ono

1　はじめに

会話は人同士のコミュニケーションにおいて最も基本的かつ日常的なものであり, 様々な場面でそれぞれに違った内容の会話が行われている. 会話の内容は, 同じ場所で同じ人物同士で話をしていても, 時間の流れと共に移り変わる. 会話に加わっている話者が積極的に発話をして盛り上がる時もあれば, 口数が少なくなり盛り上がらない時もある.

会話分析の研究分野では, 音声を精確に転写するために, 発話の重なりや沈黙, イントネーションや笑い声などの非言語情報も記述するシステムが使われている [1]. しかし盛り上がりについては, 会話分析の研究分野では明確な定義での定量的な評価は行われいない.

盛り上がりの検出は様々な角度から検討されており, ノンバーバルな音声情報や, 視線や体の動きなどの視覚情報, 生体指標などを用いて盛り上がりの自動推定に関する研究が行われている [2][3][4][5]. これらの研究では, 会話中やエージェントとのインタラクション中の盛り上がりを検出することで, アバタの表情の自動制御やエージェントによる雰囲気の判定などに役立てることを目標としており, 盛り上がりの指標はリアルタイムでの検出の可能性があるものが用いられている.

守屋らは, 会話の盛り上がりを会話活性度と定義し, 音声情報から会話活性度の自動推定を目的とし, 活性度と相関のある音声指標を分析している [3]. そこで守屋らは, 平均正規化音圧, 発話数, 総発話数, オーバーラップ数, オーバーラップ率, 平均正規化ピッチが会話の盛り上がりとの相関が高いことを示している. これらのうち, 本研究では平均正規化音圧とオーバーラップ率を盛り上がりの指標のパラメータとして用いて, リアルタイムでの盛り上がり検出に向けて研究を行う.

盛り上がりにおける重要な要素に笑いがある. 笑い声の音声認識の研究 [6] も進められているが, 盛り上がり検出の音声指標としては笑い声は扱われていない. 本研究では, 笑い声にも着目し, 盛り上がりのリアルタイム検出の指標となる可能性を検討する.

またリアルタイムでの盛り上がり検出では, 会話をされている話題ごとにどの程度盛り上がっているのかを判定できることが期待されている. しかしこれまでの先行研究では, 制約をかけた実験設定のもとでの会話や場の盛り上がりを評価しており, 会話の話題ごとの盛り上がりを評価するには至っていない. そこで本研究では, 日常的な会話から, 沈黙に着目して話題の区切れを自動検出し, それぞれの話題がリアルタイムにどの程度盛り上がったかを推定する手法を検討する.

2　盛り上がり指標の定義

本研究では, 平均正規化音圧とオーバーラップ率の平均を高揚度と定義し, 盛り上がりの指標とした. 本節では, 高揚度と笑い声についての定義を説明する.

2.1　平均正規化音圧

平均正規化音圧は, 会話の時間軸に沿って 5 秒単位で分割した区間ごとに求める. 会話全体から各話者についての音圧の平均値 $mean_A, mean_B$ と, 標準偏差 σ_A, σ_B を求める. 話者 A, B の時刻 t での音圧を V_{At}, 話者 B の音圧を V_{Bt} とすると, 各話者の時刻 t での正規化音圧 SV_{At}, SV_{Bt} は次のように定義する.

$$SV_{At} = |V_{At}| \times \frac{63}{mean_A + \sigma_A} \tag{1}$$

ただし $SV_{At} > 300$ のとき $SV_{At} = 300$ として, 上限値を 300 に設定する. SV_{Bt} についても同様に定義する. 時刻 t での正規化音圧 SV_t は,

$$SV_t = \frac{SV_{At} + SV_{Bt}}{2} \tag{2}$$

となる. 最後に, 区間 d についての平均正規化音圧の割合 SV_d を求める. 区間 d での要素数を N としたとき, 区間 d での平均正規化音圧の割合は

$$SV_d = \frac{1}{N} \times \sum_{t=1}^{N} \frac{SV_t}{300} \tag{3}$$

として求めることできる.

2.2　オーバーラップ率

オーバーラップ率も同様に 5 秒間の区間単位で求める. 区間 d で要素数が N, 時刻 t での各話者の正規化音圧が SV_{At}, SV_{Bt} として, SV_{At} と SV_{Bt} が共に閾値 20 を超えた回数を C_1 とする. 区間 d でのオーバーラップ率 O_d は次のように定義する.

$$O_d = \frac{C_1}{N} \tag{4}$$

北海道大学大学院情報科学研究科, Graduate School of Information Science and Technology, Hokkaido University (†)

2.3 高揚度

高揚度は本研究での盛り上がりの指標として用いる．高揚度 T_d は区間 d での平均正規化音圧の割合とオーバーラップ率の平均として定義する．

$$T_d = \frac{SV_d + O_d}{2} \tag{5}$$

2.4 笑い声

笑い声の種類は, 快 (自発的な喜び) の笑いと社交的（あいづちやうなずき）な笑いとに大別される [7]．本研究での笑い声の定義は, ハハハハと表記されるような「は」や「ふ」などの音で構成されて, 明確に聞き取れるものを笑い声とした．すなわち, 喋りながらの笑い声や, 引き笑いや鼻笑いなどは除外し, 検出の対象とはしなかった．笑い声の認識は, HTK[8] を用いて学習モデル HMM を作成し, 学習データとして本研究で定義した笑い声を与えて実現させた．特徴量は MFCC を用いた．

3 話の区切れと沈黙

会話や場の盛り上がりをリアルタイム検出する場合, 各時刻での盛り上がりの検出と共に話題ごとの盛り上がりの検出が, 会話や場の雰囲気の判定に役立つと考えられる．そこで本研究では話題転換時の特徴の一つである沈黙 [9] に着目して, 会話の話題の区切れを検出する方法を検討する．沈黙の指標として, 本研究では沈黙率を定義する．沈黙率も 5 秒間の区間単位で求める．区間 d で要素数が N, 時刻 t での各話者の正規化音圧が SV_{At}, SV_{Bt} として, SV_{At} と SV_{Bt} が共に閾値 20 以下の回数を C_2 とする．区間 d での沈黙率 S_d は次のように定義する．

$$S_d = \frac{C_2}{N} \tag{6}$$

沈黙率が $S_d >= 0.65$ の時, 話題の区切れとして定義した．

4 システムの構成

本節では, 本研究で盛り上がりを検出するために作成したシステムについて説明する．盛り上がり検出システムは, 同期させた二つの音声ファイル (WAVE 形式) を入力として与えると, 盛り上がりに関するパラメータを各区間ごとに出力し, CSV 形式ファイルとして保存する．システムは, 横軸に時間軸 [s], 縦軸に高揚度 [%] の軸として, 5 秒ごとに高揚度と沈黙率を表示させるようにした (図 1)．沈黙率によって話題の区切れを定め, 高揚度によって盛り上がりの程度を計測した．また笑い声については, 隠れマルコフモデルを用いたシステムによる音声認識で検出した．

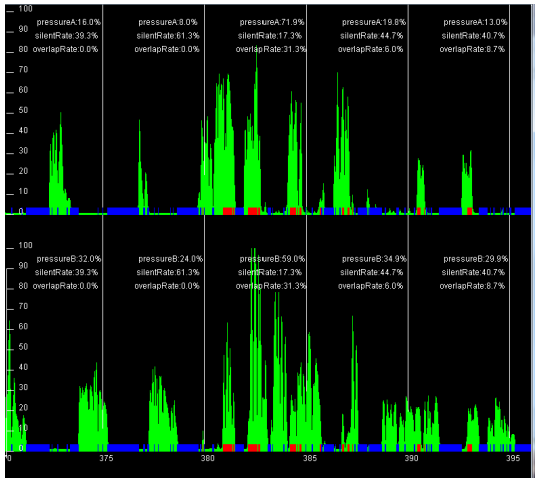

図 1: 本システムの実行画面

5 実験

5.1 実験設定

実験は, 1 対 1 の電話による会話を 2 組計 4 人で行った．実験時間は 10 分程度で, 被験者に対して会話の内容などの制約をかけなかった．実験終了後, 被験者にそれぞれアンケートによる調査を行った．アンケートは, 実験中に録音しておいた被験者自身の会話音声を聞きながら, 時間軸に沿って 5 段階の盛り上がりの主観評価を行わせた．

5.2 実験結果 1

1 組目の実験結果を図 2 に, 2 組目の実験結果を図 3 に載せる．アンケート評価は二人の被験者の評価の平均である．3 節で定義したように, 沈黙率が 65% 以上の時点でシステムは話題の転換点と認識し, 図 2 の会話では 4 つの話題に, 図 3 の会話では 8 つの話題に分割された．2 組で計 12 の話題について, それぞれの話題での平均高揚度と平均アンケート評価についての相関係数を求めると R = 0.875 となり, 高い相関が得られた．

5.3 実験結果１の考察

沈黙率は 5 秒間という短い区間での沈黙の割合なので, リアルタイムでも話題の区切れの検出が実現可能である．それぞれの話題についての高揚度の平均とアンケート評価の平均の相関も高いので, 沈黙率による話題の区切れの検出方法は, 話題ごとの盛り上がり検出の一つの手法としての可能性を示せた．しかし, 今回の実験でシステムが検出した話題の区切れ以外にも, 多くの話題の区切れが会話中には存在するので, 今後も調査を進めていく必要がある．

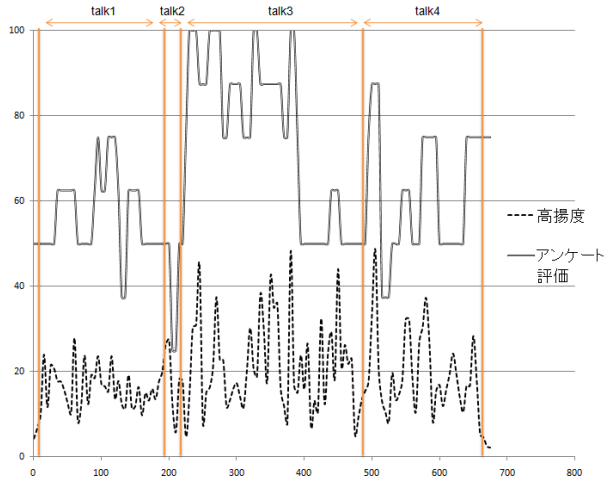

図 2: 1 組目の実験結果

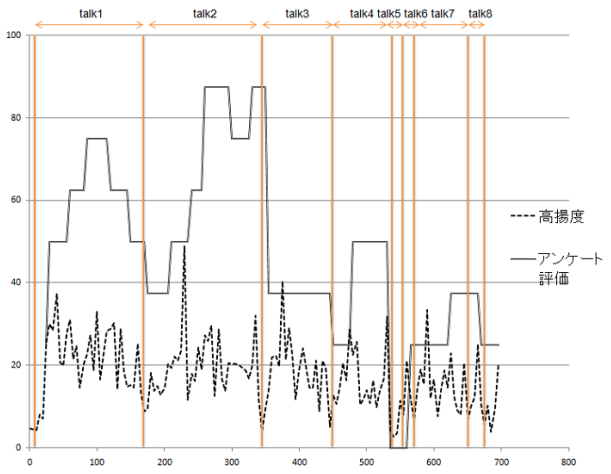

図 3: 2 組目の実験結果

5.4 実験結果 2

図 2, 図 3 に笑い声が生じた時点にマーキングしたものを図 4, 図 5 に示す. 1 組目の会話では, 笑い声は 22 回生じた. 笑いが生じた時点の時刻とその時の高揚度を表 1 に示す. 笑い声が生じた時点での高揚度の平均は 30.17 %で, 会話全体の高揚度の平均 18.80 %を大きく上回った. しかし, 笑い声が生じた時点の高揚度が, 会話全体の高揚度の平均を下回る場合も確認された. 2 組目の会話は, 笑い声が 6 回生じた（表 2）. 2 組目の笑い声が生じた時点での高揚度の平均は 30.19 %で, こちらも会話全体の高揚度の平均 17.77 %を大きく上回った.

5.5 実験結果 2 の考察

笑い声が生じている所は, 高揚度が高く, 盛り上がりとの相関が見られるが, 例外として, 高揚度が高くない所で笑い声が生じている場合があった. これは, 笑い声の種類として, 高揚度の高い所での笑いは快の笑い, 高揚度の低い所での笑いは社交的な笑いの可能性が高いと考えられる. 盛り上がり検出における笑い声認識は, 高揚度が高くて笑い声が生じている時点, つまり喜びを伴う快の笑い声が生じた時点を検出できる可能性があることを確認できた.

表 1: 1 組目の笑い声生起時刻と高揚度

時刻 [s]	高揚度 [%]	時刻 [s]	高揚度 [%]
97	23.68	351	42.33
123	13.63	365	36.13
152	11.93	381	48.38
216	18.3	382	48.38
232	14.18	409	26.55
236	30.45	452	44.13
238	30.45	502	33.2
272	37.55	505	48.62
275	22.88	558	32.35
278	22.88	578	30.42
330	18.95	650	28.4

表 2: 2 組目の笑い声生起時刻と高揚度

時刻 [s]	高揚度 [%]
64	31.05
232	48.95
262	27.325
489	25.7
499	11.725
530	31.88

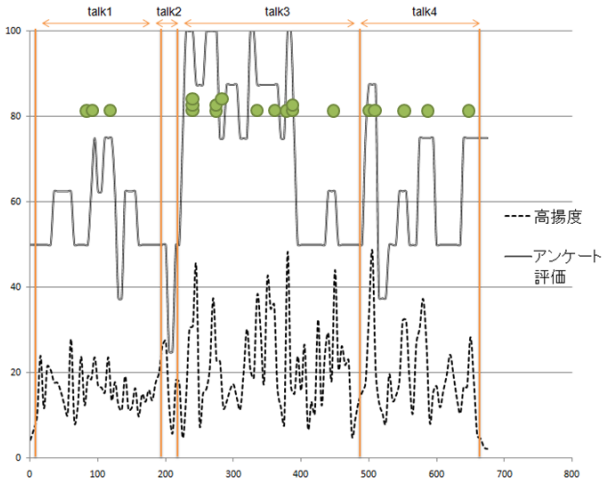

図 4: 1 組目の実験結果・笑い声マーキング

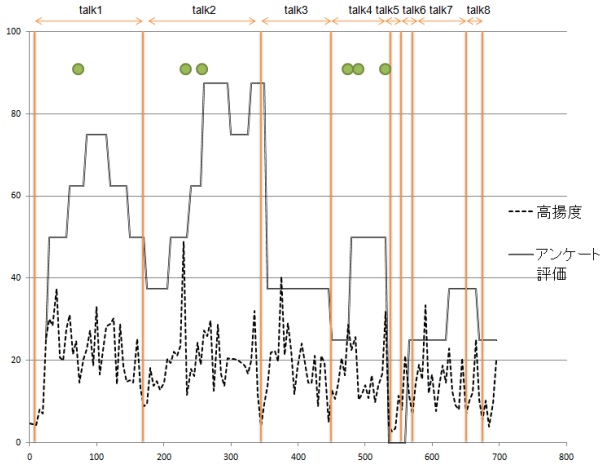

図 5: 2 組目の実験結果・笑い声マーキング

6 まとめと今後の課題

本研究では, リアルタイムでの盛り上がりの検出に向けて, 沈黙率を用いた話題の区切れの自動推定を行い, 各話題での盛り上がり検出を試みた. その結果, 各話題で高揚度と被験者によるアンケート評価との相関が高いという事を示した. また盛り上がり検出の音声指標として笑い声に着目して, 笑い声が生じる所は高揚度も高いということを示し, 笑い声の盛り上がりにおける指標としての可能性を見出した. 逆に高揚度の高さによって, 生じた笑い声が快の笑いなのか社交的な笑いなのかを判別する可能性も示した.

今後の課題としては, 多くの実験を重ねて, アプリケーションを視野に入れた盛り上がり検出のパラメータ調整を行う必要がある. 本研究では高揚度と沈黙率を用いて話題ごとの盛り上がりの検出の可能性を示したが, よりリアルタイム性が求められるアプリケーションを想定した場合, 本研究の結果だけでは不十分である. 今回は 5 秒間という区間の平均高揚度とアンケート評価との比較を行ったが, よりリアルタイム性が求められるアプリケーションでは 5 秒という区間は長い. またアンケートでの厳密な時刻と対応した盛り上がり評価は困難であるため, リアルタイム性を考慮すると生体信号などを用いた評価を検討する必要がある.

参考文献

[1] H. サックス, E.A. ジェグロフ, G. ジェファソン : ”会話分析基本論集-順番交代と修復の組織”, 世界思想社.

[2] 宮島 俊光, 藤田 欣也 : ”音声チャットシステムにおける基本周波数と音圧を利用したアバタ表情制御法”, ヒューマンインタフェース学会論文誌 Vol.9, No.4, 2007.

[3] 守屋 悠里英, 田中 貴紘, 宮島 俊光, 藤田 欣也 : ”ボイスチャット中の音声情報に基づく会話活性度推定可能性の検討”, ヒューマンインタフェース学会研究報告集 Vol.13 No.3.

[4] 大本 義正, 三宅 峰, 西田 豊明 : ”複数ユーザインタラクションにおける外発的な盛り上がりの雰囲気の判定方法と影響の検討”, 電子情報通信学会論文誌, D Vol.J93-D, No6, pp.870-878, 2010.

[5] 前田 貴司, 高嶋 和毅, 梶村 康祐, 山口 徳朗, 北村 喜文, 岸野文朗, 前田 奈穂, 大坊 郁夫, 林 良彦 : ”3 人会話における非言語情報と「場の活性度」に関する検討”, 電子情報通信学会技術研究報告. ヒューマンコミュニケーション基礎 109(457), 73-78, 2010-03-01.

[6] M. Knox, N.Morgan and N. Mirghafor : ”Getting the Last Laugh: Automatic Laughter Segmentation in Meetings”, in Interspeech, 2007.

[7] 清水 彰, 角辻 豊, 中村 真 : ”人はなぜ笑うのか”, 講談社, 1994-6.

[8] Hidden Markov Model Toolkit : http://htk.eng.cam.ac.uk/

[9] 楊　虹 : ”中日母語場面の話題転換の比較 -話題終了のプロセスに着目して-” , 世界の日本語教育 2007 年 6 月 17 日.