自然な対話を実現する合成音声デザインの一考察

関根雅人[†] 小川克彦[††] 野本済央[†††]
井島勇祐[†††] 吉岡理[†††]

本研究では、合成音声を用いた対話コンテンツデザインへの応用を想定し、合成音声同士の対話における先行話者と後続話者の抑揚及び話速の関係が、同意発話の自然さに与える影響を考察した．抑揚・話速、それぞれについて、先行話者と後続話者との差が異なる数種類の対話セットを用意し、ランダムに抽出された二つの対話セットを被験者に提示し、同意の発話としてより自然な方を選択してもらう官能評価実験を行った．実験結果の分析から、抑揚・話速ともに、先行話者の韻律パラメータによって、自然な同意の発話として評価される後続話者音声の傾向が変ることが考察された．

A design of naturally spoken dialogue by Speech Synthesis System

Masato SEKINE,[†] Katsuhiko OGAWA,[††]
Narichika NOMOTO,[†††] Yusuke IJIMA[†††]
and Osamu YOSHIOKA[†††]

In this research, as a future application of dialogue design using speech synthesis system, we study the relationship between the intonation and speech rate of a dialogue between 2 speakers, and the nature of an agreeing conversation. We prepared several sets of dialogue that differ each in intonation, and speech rate, from the former speaker and the subsequent speaker. Presenting 2 randomly selected sets of dialogue, we held a sensory evaluation experiment, asking subjects to select the set of dialogue which sounds more like a natural conversation of agreement.Analyzing results of the experiment, we have recognized that a tendency in which subjects perceive dialogue as a natural agreeing conversation varies depending on a former speaker's prosody parameters.

1. はじめに

近年、テレビ CM や商品販促サイト内のムービーにおいて、出演者同士がフリートーク形式で商品のポジティブな印象を語り合う演出が盛んに用いられ始めている．出演者同士の自然な”おしゃべり”の中で商品が好意的に扱われるシーンに影響され、視聴者も商品に対して好意的な印象を持つことを狙う演出であると考えられる．こうしたプロモーショントークの演出は、好意的な印象を自然に発話することのできる話者を複数人要し、プロモーション対象が変わるごとに収録をし直す必要があるなどの制作コストが生じる．

テレビ番組や e-ラーニングビデオのナレーションなど、これまで声優を要する収録コストの生じていたコンテンツ制作の場に合成音声が用いられ始めているが、人間同士の円滑で自然なコミュニケーションを、機械同士の音声対話で模す事ができれば、ナレーション利用や音声応答だけでなく、上述のようなトーク・コンテンツ制作への応用が可能になると考えられる．そのためには、従来の音声研究の中心的な分析対象であった単一話者の音声情報だけでなく、対話における話者間での音声情報の相互影響についても考慮する必要がある 1)．

2. 本研究の位置付けと目的

対話における音声情報の相互影響は、人間の発話に対し自然に応答する音声対話システムの研究において注目され始めている．東海林らの研究 2) では、対話リズム（発話内の小休止）の同調をおこなう音声対話システムが提案されている．リズムの同調以外にも音声の韻律（感情を含む）などの応答能力を人間に近

† 慶應義塾大学大学院 政策・メディア研究科
Keio Universigy Graduate School of Media and Governance
†† 慶應義塾大学 環境情報学部
Keio Universigy Faculty of Environment and Infomation Studies
††† NTT サイバースペース研究所　NTT Cyber Space Laboratories

づける事が、自然な対話を促す為に有効であることも言及されている．

従来研究では、ユーザフレンドリーな音声応答システムへの応用が想定されいたため、＜人― 機械＞の二者間の対話コミュニケーションの上で、ユーザーである人間は対話相手としての機械の音声を判断する．本研究で扱うコミュニケーションの図式は、＜＜機械 － 機械＞ ― ＜それを聴く人間＞＞であり、人間は、対話に直接関わらない第三者として機械の音声を判断する．従って、＜人－機械＞の場合とは機械の音声を聴く態度が異なってくると考えられる．また、本研究ではプロモーション・トークでの応用を想定しているが、こうしたプロモーション・トークでは、ポジティブな発話に対する自然な同意の意図の表現が重要になってくる．韻律による「意図」の表現は、従来研究では感情音声表現の中に含まれてきたが、3) のプラグマティック・イントネーション理論では、韻律情報に階層を設定することで、emotion（感情）の表現と、attitude (態度) とが区別されている．前者が独立で表現が可能であるのに対し、態度（意図）は多くの場合対話相手がいることが前提となるため、対話を聴く第三者が対話中のある発言に対してその意図の判断をする際に、発言そのものに含まれる韻律情報だけではなく、その前後の発言の韻律についても意図の判断材料として利用している可能性が考えられる．本研究では、このような視点から、対話中の韻律の相互関係に着目し、合成音声同士の対話音声の聴取実験を通じて、先行話者と後続話者との韻律情報の相互関係と、ポジティブな発話に対する同意発話の自然性との関係について傾向を考察する．

3. 実 験 内 容

先行話者と後続話者の話し方の関係性を調査するために、被験者に韻律情報の異なる複数の合成音声対話セットを聴いてもらい、先行話者に対する後続話者の応答が、同意の示しかたとして自然かどうかを調べる官能評価実験を行った．実験は大学生 33 名を被験者に行った．

3.1 対 話 内 容

対話内容は、先行話者（A さん）が架空の商品「ナペタ」に対してポジティブな発話を行い、それに対して、後続話者（B さん）が応答する．このとき、B さんの発話内容は、意味論的に特定の意図を含まず、韻律によって意図が左右される台詞を設定する．

対話スクリプト）

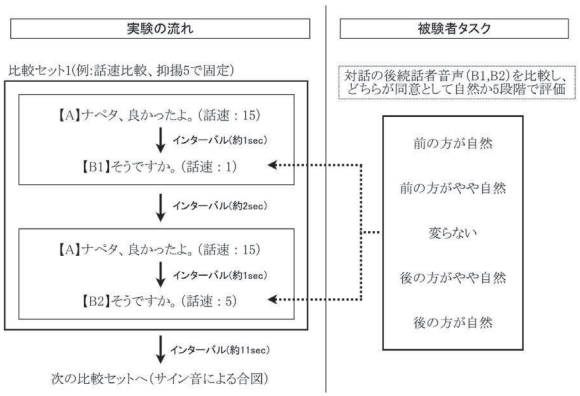

図 1 実験の流れ及び被験者タスク
Fig. 1 flow chart of the test and a task of subjects

A さん）ナペタ、良かったよ．
B さん）そうですか．

3.2 音声サンプルの作成

NTT サイバースペース研究所で開発した音声デザインツール「Sesign」4) を使用し、女性（A さん）、男性（B さん）の２種の話者モデルによる音声サンプルを作成した．話速と抑揚（ピッチのダイナミックレンジ）のパラメータを変更する事で、それぞれ複数種の音声ファイルを制作した．(今回使用した音声ファイルの詳細については、実験結果及び分析の章の**表 1 表 2** 内に記載．)

3.3 提示方法と評価基準

今回の実験では、話者同士の話速の相互関係と、抑揚の相互関係について調べるため、

(1) 話速を固定し（先行話者話速≒後続話者話速）抑揚の異なる二つの後続話者の音声を比較する対話セット (8 種)

(2) 抑揚を固定し（先行話者抑揚 ≒ 後続話者抑揚）話速の異なる二つの後続話者の音声を比較する対話セット (7 種)

(3) 抑揚・話速が共に異なるダミーの対話セット (5 種)

三つのカテゴリを用意し、合計 20 セットをランダムな順番で提示した．実験の流れ、被験者タスクを**図 1** に示す．音声は PC 上で再生したものを外付けのスピーカーから出力し、一回の実験で複数人（10 人程度）の被験者に提示している．

4. 実験結果及び分析

4.1 結 果 集 計

被験者に各対話セット毎にどちらが自然に聞こえるかを 5 段階で評価してもらったアンケート結果を集計

し得点を算出した．得点は、＜前の方が自然＞および＜後ろの方が自然＞の場合、選ばれた方に 1 点加算、＜前の方がやや自然＞及び＜後の方がやや自然＞の場合、選ばれた方に 0.5 点加算、＜変らない＞が選択された場合はどちらの対話にも得点を入れない．それぞれの対話について、全ての被験者の得点を合計したものを被験者人数で割ることで平均得点を算出した．

4.2 音声データの解析

実験で使用した音声サンプルの韻律情報は、音声解析ソフト「praat」5) を用い、praat 上で Yi Xu が制作した praat script である「ProsodyPro」6) を使用し解析を行った．ここでは、各音声サンプルについて、平均モーラ長、各モーラ内での平均基底周波数 (meanF0) データを得た．モーラの切り分けについては、音声を聞きながら波形の目視によって行った．さらに、全音声サンプルについて、平均基底周波数データから最少基底周波数 (minF0) と最大基底周波数 (maxF0) を求めた．

4.3 抑揚の影響について

話速を固定し、抑揚のみを変化させた対話セットの比較実験結果から、抑揚の相互関係が同意発話の自然さの評価に与える影響の分析を行った．女性話者モデルと男性話者モデルとでは、抑揚変化の絶対量が異なる為、分析には F0 の変化量を用いるのではなく、Sesign の抑揚パラメータを用いることにした．実際の F0 の値は表 1 に付記する．それぞれの話者モデルのおおよその抑揚変化特性について、パラメータの値を P とすると、女性話者モデルでは

$$maxF0 \simeq 194.4[Hz] + 8.1[Hz] * P$$
$$minF0 \simeq 207.1[Hz] - 2.7[Hz] * P$$

になっており、男性話者モデルは、

$$maxF0 \simeq 97.5[Hz] + 5.1[Hz] * P$$
$$minF0 \simeq 91.0[Hz] + 3.1[Hz] * P$$

になっている．女性話者モデルの場合には $minF0$ と抑揚パラメータは反比例関係にあるが、男性話者モデルの場合、$minF0$ は抑揚パラメータと比例しており、抑揚を大きくすると全体的に F0 が上にシフトする特性がある**図 2**.

図 3 は、表 1 の話者間の抑揚差と得点の関係を散布図に示したものであるが、先行話者 A の抑揚ごとの傾向を視る為に、A の抑揚で系列を分けて表示している．この図から以下の傾向が読み取ることができる．

- 後続話者 B の抑揚、速度のパラメータがどちらも 1 の場合 (表 1 参照)、先行話者のパラメータに関わり無く、不自然な発話と判断される傾向がある．

表 1 抑揚比較対話セットの結果
Table 1 results of intonation comparison experiments

合成パラメータ	抑揚差 (B−A)	A:F0レンジ	B:F0レンジ	平均得点
A(15,5) B1(15,20)	15	196.6–233.6	160.9–207.3	0.152
A(1,5) B1(1,1)	−4	196.0–232.0	114.8–119.7	0.015
A(1,5) B1(1,20)	15	196.6–233.5	158.4–207.9	0.379
A(15,5) B1(15, 10)	5	196.0–232.0	134.7–144.7	0.318
A(15,5) B2(15,10)	5	196.6–233.6	207.3–145.6	0.364
A(1,5) B2(1,5)	0	196.0–232.0	108.5–127.8	0.212
A(1,5) B2(1,1)	−4	196.6–233.5	114.8–119.7	0.106
A(15,5) B2(15, 5)	0	196.0–232.0	113.5–127.7	0.121
A(10,10) B1(10, 20)	10	178.3–275.4	163.8–209.8	0.242
A(10,10) B1(10,20)	10	178.3–275.4	163.8–209.8	0.182
A(10,10) B2(10,5)	−5	178.3–275.4	113.3–127.3	0.288
A(10,10) B2(10,10)	0	178.3–275.4	110.5–143.8	0.379
A(1,20) B1(1,1)	−19	153.9–362.7	114.8–119.7	0.121
A(1,20) B1(1,10)	−10	153.9–362.7	112.0–144.3	0.167
A(1,20) B2(1,10)	−10	153.9–362.7	112.0–144.3	0.44
A(1,20) B2(1,20)	0	153.9–362.7	158.4–207.9	0.41

※合成パラメータは、話者（話速設定値，抑揚設定値）．
※抑揚差は(B:抑揚設定値) − (A:抑揚設定値)で算出．
※F0レンジは、minF0−maxF0．

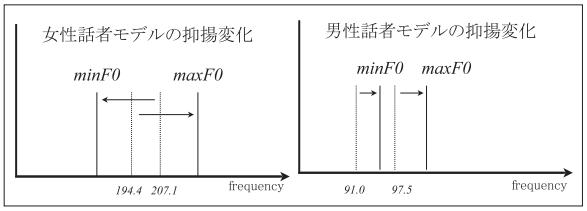

図 2 Sesign 話者モデルの抑揚変化特性
Fig. 2 intonation-change characteristic of each speech model in Sesign.

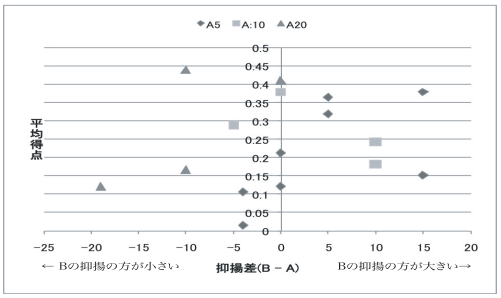

図 3 抑揚差と得点の関係
Fig. 3 relation between differences of intonation and score .

- 先行話者の抑揚が小さい場合（A5 の系列)、後続話者の抑揚はやや大きめのときが自然な同意発話として判断される傾向がある．
- 先行話者の抑揚が中程度の場合（A10 の系列)、後続話者の抑揚は同程度のときが自然な同意発話として判断される傾向がある．

- 先行話者の抑揚が大きい場合（A20 の系列)、後続話者の抑揚は同程度かやや小さめのときが自然な同意発話として判断される傾向がある．

4.4 話速の影響について

抑揚を固定し、話速のみを変化させた対話セットの比較実験結果から、話速の相互関係が同意発話の自然さの評価に与える影響の分析を行った．Sesign では、話速については話者モデルに関わり無く変化量が一定なため、分析では音声データの解析で得た平均モーラ長を用いる．各対話の結果を表 2 に示す．

表 2 話速比較対話セットの結果
Table 2 result of speech-rate comparison experiments

合成パラメータ	話速差 B−A	A平均モーラ長	B平均モーラ長	平均得点
A(1,5) B1(5,5)	−31.71	246.84	215.12	0.030
A(1,5) B2(1,5)	1.70	246.84	284.54	0.121
A(1,20) B1(10,20)	−76.36	232.03	155.67	0.455
A(1,20) B1(1,20)	2.32	232.03	234.35	0.106
A(1,20) B2(20,20)	−154.43	232.03	77.60	0.121
A(1,20) B2(10,20)	−76.36	232.03	155.67	0.333
A(10,10) B2(10,10)	−0.06	168.85	168.79	0.182
A(10,10) B1(5,10)	43.21	168.85	212.06	0.424
A(10,10) B1(15, 10)	−42.12	168.85	126.73	0.182
A(10,10) B2(20, 10)	−87.26	168.85	81.59.59	0.136
A(15,5) B1(1,5)	119.81	128.73	248.54	0.121
A(15,5) B1(15,5)	−2.33	128.73	126.40	0.485
A(15,5) B2(20,5)	−44.36	128.73	84.36	0.303
A(15,5) B2(20,5)	−44.36	128.73	84.36	0.061

※合成パラメータは、話者（話速設定値, 抑揚設定値).
※話速差は(B平均モーラ長) −(A平均モーラ長)で算出.
※話速差、A平均モーラ長、B平均モーラ長の単位はmsec.

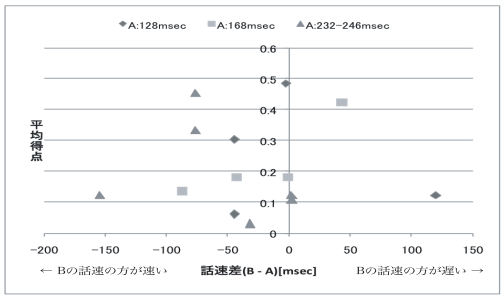

図 4 話速差と得点の関係
Fig. 4 relation between differences of speech-rate and score

また、**図 4** は、表 2 の話者間の話速差と得点の関係を散布図に示したものであるが、先行話者 A の話速ごとの傾向を視る為に、A の話速で系列を分けて表示している．この図から以下の傾向が読み取ることができる．

- 先行話者の話速が速い場合（A128msec の系列)、後続話者の話速は同程度のときが自然な同意発話として判断される傾向がある．
- 先行話者の話速が中程度の場合（A168msec の系列)、後続話者の話速はやや遅めのときが自然な同意発話として判断される傾向がある．
- 先行話者の話速が遅い場合（A232 246msec の系列)、後続話者はやや早めのときが自然な同意発話として判断される傾向がある．

5. む す び

合成音声対話の聴取実験を通じ、抑揚・話速、それぞれについて、先行話者の韻律情報ごとに、自然な同意の発話として判断される傾向が変わってくることが考察された．今後の課題としては、今回の実験では一種類の対話スクリプトで実験を行ったため、他の台詞でも今回の様な傾向が再現されるか実験する必要がある．また、抑揚比較、話速比較の聴取実験を同時に行ったため、刺戟数が少なく、詳細な傾向までは得ることができなかった．今後はそれぞれの場合についてより詳細な実験を行いつつ、話速と抑揚の相互作用についても調査し、最終的にこれらの傾向のモデル化を検討して行く．

参 考 文 献

1) 長岡千賀ほか：音声対話における交替潜時が対人認知に及ぼす影響, ヒューマンインターフェースシンポジウム 2002, 一般発表 (2002).
2) 東海林圭輔ほか：対話に関するリズムや同調作用を考慮した音声対話システム, 情報処理学会 研究報告, Vol.2006 No.40, pp.43-48(2006).
3) ニックキャンベル：プラグマティック・イントネーション：韻律情報の機能的役割, 文法と音声, 黒潮出版 pp. 55-74(1997)
4) 阿部匡伸ほか：音声デザインツール Sesign, 電子情報通信学会論文誌 Vol. J84-D-Ⅱ No.6 pp. 927-935(2001)
5) Boersma, Paul & Weenink, David (2010). Praat: doing phonetics by computer[Computer program]. Version 5.2.07, retrieved 24 December 2010 from http://www.praat.org
6) Xu, Y. (2005-2010). ProsodyPro.praat. Available from: http://www.phon.ucl.ac.uk/home / yi /ProsodyPro/.