J-012

合成音声を利用した好印象発話モデルの構築

Developing Good Impressions Speech Model Using Synthesized Speech

浅田 龍星 † 西田 悠 † 中平 勝子 † 北島 宗雄 †
Ryusei Asada Haruka Nishida Katsuko T. Nakahira Muneo Kitajima

1 はじめに

コミュニケーションは，日常生活に必要不可欠なものである．コミュニケーションをとる主な手段は対話であり，対話によってその人の印象が決まる．対話に必要な主たる要素は「聴く」「発話する」の 2 種であるが，特に「発話する」という行為は自身の考えを相手に対して齟齬なく伝えるという意味で重要である．対話相手に与える自身の印象は，特に良い印象を与えたいと自身が感じる場面，例えば面接の場における自己アピールや，相手に伝わりやすいプレゼンテーションなど．本稿では，発話印象のなかでも，特に「良い」印象，すなわち，好発話印象の構成要素とその構造を呈示し，円滑なコミュニケーションとは何か，の考察に寄与する．

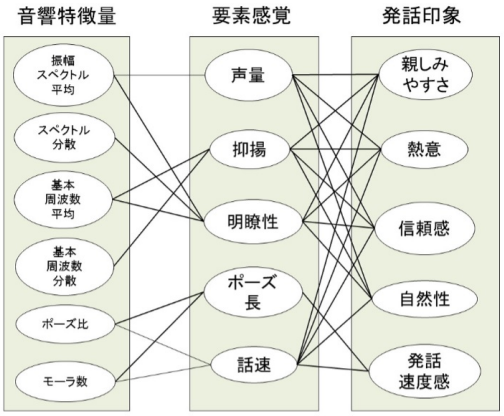

図 1 発話印象のモデル．

2 発話印象モデル

発話印象モデルは 6 つの音響特徴量（振幅スペクトル平均・スペクトル分散・基本周波数平均・基本周波数分散・ポーズ比・モーラ数），5 つの要素感覚（声量・抑揚・明瞭性・ポーズ長・話速），5 つの発話印象（親しみやすさ・熱意・信頼感・自然性・発話速度感）をある一定の関係で接続することによって記述することができる [1]．このうち，音響特徴量と要素感覚の関係分析およびその指標化関しては西田ら [2] が着手しているため，本稿では要素感覚，発話印象についての関係を記述するものとする．図 1 に発話印象のモデルを示す．

要素感覚とは音声を聴いた際に知覚されるものを感覚的に評価したものであり，発話印象は音声の感覚的評定から認知される話し方についての印象評価である [1]．籠宮ら [3] の研究により発話速度感とポーズ比，モーラ数の関係性分析が行われており，内田ら [4] の研究では抑揚と自然性の関係性分析が行われている．西崎ら [5] らの研究では熱意と声量，抑揚，明瞭性に関連があることが明らかとなっているため，発話印象の項目に親しみやすさ，熱意，信頼感，自然性，発話速度感を設定した．また，今回の実験では，各要素感覚が与える満足度の影響を調べるため，全体の満足度といった項目を追加した．このモデルをもとに，要素感覚の組み合わせの変化による発話印象の感じ方の違いを得ることにより，どのような要素感覚を含む発話でどのような印象を与えることができるか，ということが明らかになると考えられる．そこで，好印象発話モデルを構築するために，様々な要素感覚を含む合成音声を被験者に聞かせ，発話印象の評価を行わせるという印象評価実験を行う．

3 実験

本章では，発話音声を特徴付ける要素感覚より引き起こされる，好印象発話モデルの構築を目的として，次の実験を行う．肉声ではなく合成音声を用いる理由として，肉声では求められる様々なパターンの要素感覚を正確に再現することが困難な点，普段聞きなれない合成音声を用いることでバイアスがかからず，フラットな評価を期待できる点，要素感覚のパラメータを明確な値に基づき設定可能な点が挙げられる．

合成音声を作成するにあたり，比較的操作がしやすい「棒読みちゃん」というフリーソフトを用いる．用いる要素感覚に関して，棒読みちゃんにおいて設定すること

† 長岡技術科学大学

が可能な「声量・抑揚・話速」の 3 つに絞り合成音声を作成する．これらの 3 要素を大小の 2 段階に分けるため，実験に用いる合成音声は合計 8 パターンとなる．声量はデシベルの大きさによって大まかにどの程度のうるささを持つか決まっているため，それを参考に実験で用いる大きさを決定した．

モデルを構築する際に必要となる，要素感覚の組み合わせの変化に伴う発話印象の評価の変化を数値として測定し，分析を行う．また，分析を行い得られた結果をもとにモデルの構築を行う．実験対象者は，聴覚に問題のない，22~45 歳までの男女 9 人とする．音声評価実験を行うにあたり，防音加工が施されているスタジオにて実験を行った．

音声のパターンは以下の表のようになる．表 1 に実験で用いる要素感覚のパターンを示す．

表 1　要素感覚のパターン

パターン	声量	抑揚	話速
A	大	大	大
B	大	大	小
C	大	小	大
D	大	小	小
E	小	大	大
F	小	大	小
G	小	小	大
H	小	小	小

実験に用いる文章は，青空文庫に所蔵される作品から無作為に 8 つ選定し，セリフを含まない，地の文のみで構成された部分を抜粋する．また，文章の内容に関係なく合成音声の要素感覚のみでの評価を得たいため，用いる文章は引き込まれやすい物語の導入部ではなく，物語の途中から抜粋した．

要素感覚に関して，実験に用いる音声の大きさの目安を表 2 に示す．これらを踏まえて，今回の実験では声量大を 100db，声量小を 30db として設定した．

抑揚に関しては明確な値を設定することはできないが，用いる合成音声により強弱をつけることが可能だと判断した．棒読みちゃんでは男性の声，女性の声，中性的な声，機械的な声，女性の声とは別に “Haruka” という声が使用可能となっている．抑揚が大きい音声には Haruka を用い，抑揚が小さい音声には中性の合成音声を用いた．Haruka が女性の声のため，抑揚が小さい音声も女性の声で統一しようと考えたが，残る女性の合成音声はインターネット上の動画などでよく使われるものであり，聞く機会が比較的多いことから評価項目にバイアスがかかってしまう可能性を考慮し，女性の声に近く，聞く機会も少ないであろう中性の合成音声を抑揚が小さいほうの音声に採用した．

話速に関しては，NHK のニュースキャスターは 1 分間に平均 300 文字を朗読するとされている [8]．また，高齢者にとっては通常のニュースを読み上げる速度を 0.8 倍にしたほうが読みやすいという先行研究も存在する [8]．そこで，それらを目安に，読み上げる文章を全て 300 文字に統一し，話速小を 1 分間の 1.25 倍の速さ，話速大を 1 分間の 0.8 倍の速さとなるように設定した．

合成音声の再生順序は乱数の生成により事前に決める．要素感覚の組み合わせが 8 通りのため，再生順序も 8 通りとなる．実験の際は各 1 分ほどの合成音声をスピーカーにより再生し，発話印象のモデルにあった 5 つの発話印象（親しみやすさ・熱意・信頼感・自然性・発話速度感）に全体の満足度という項目を加え，合計 6 項目を評価してもらう．評価はアンケート用紙に記入する形式を取り，6 件法で回答してもらった．一度に行える最大実験人数は 2 人までとし，スピーカーと被験者の間は 100cm，被験者同士の間は 75cm の距離を取った．

4　実験結果

実験により得られた要素感覚の平均値および分散を表 3 に示す．また，評価値のレーダーチャートを図 2，誤差範囲を図 3 に示す．以下に，要素感覚のパターンと発話印象の評価値について，考察した結果を示す．

親しみやすさの項目には目立った特徴が見られないが，要素感覚 3 つのうち 2 つの項目が小のレベルになっ

表 2　音声の大きさの目安

最小値 ~ 最大値(db)	大きさの目安	実例
20 ~ 30	非常に静か	ささやき声，深夜の郊外
40 ~ 50	普通	図書館内，静かな事務所内
60 ~ 70	少しうるさい	洗濯機，トイレの洗浄音，テレビ
80 ~ 90	かなりうるさい	カラオケ，電車内，工事現場
100 ~ 120	聴覚機能に異常をきたす	自動車のクラクション，ジェットエンジン

表 3　発話印象の平均値と分散

声量-抑揚-話速	親しみやすさ	熱意	信頼感	自然性	発話速度感	満足度
パターン A 大-大-大	3.22 (1.44)	3.22 (1.69)	3.11 (0.86)	2.78 (1.19)	4.00 (1.00)	3.22 (0.69)
パターン B 大-大-小	3.00 (1.00)	2.56 (1.28)	2.89 (2.61)	2.78 (0.94)	2.56 (1.52)	3.00 (0.50)
パターン C 大-小-大	3.11 (0.86)	3.22 (1.94)	2.89 (1.36)	3.00 (1.50)	4.00 (0.75)	3.44 (1.28)
パターン D 大-小-小	2.11 (1.36)	2.56 (1.03)	2.67 (1.00)	2.33 (1.25)	2.89 (0.61)	2.67 (1.00)
パターン E 小-大-大	3.67 (1.25)	2.89 (1.61)	3.67 (1.75)	3.89 (1.86)	4.33 (0.75)	3.89 (0.86)
パターン F 小-大-小	2.67 (0.75)	2.33 (1.00)	2.67 (2.00)	2.78 (0.69)	3.00 (1.00)	3.00 (1.00)
パターン G 小-小-大	2.89 (1.61)	2.89 (1.36)	3.00 (2.00)	3.00 (2.50)	4.00 (1.00)	3.22 (0.94)
パターン H 小-小-小	3.11 (0.86)	2.44 (1.53)	2.89 (1.36)	2.89 (1.61)	3.56 (0.52)	3.33 (0.50)

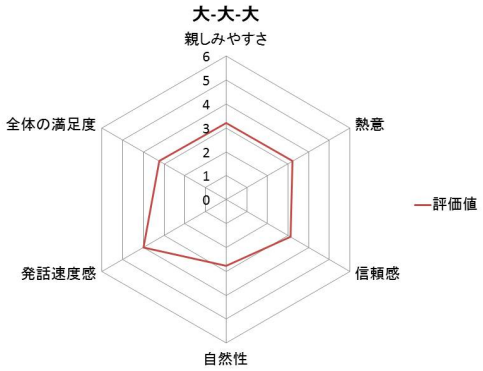

図 2　パターン A の平均値.

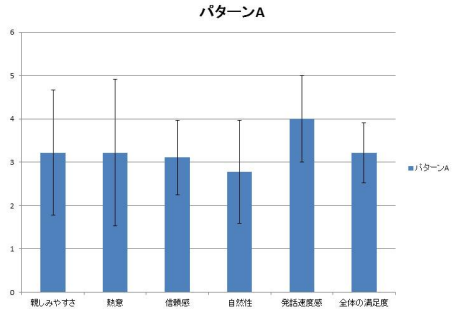

図 3　パターン A の誤差範囲.

ている，すなわちパターン D，F，G のとき，2.11，2.67，2.89 と比較的評価値が低くなっており，分散は 1.36，0.75，1.61 とそれほど高くはない．要素感覚が全て小のレベルになっているときは 3.11 と高めの評価を得ているため，要素感覚が小さくかつ全てが小ではない場合に親しみやすさをあまり感じないと考えられる．

熱意の項目は，声量，話速がともに大のときにともに 3.22 と好評価を得ている．評価が最も低いのは，声量，話速がともに小になっているときの 2.33 であるため，熱意を感じる音声は声量，話速がともに大きいものであると言える．抑揚が大のときと小のときで評価値が変わっていないのは，おそらく抑揚大小の両方ともにどこかイントネーションのおかしい部分があり，抑揚にあまり差を感じなかったためだと思われる．

信頼感に関しては，話速が大のときに評価値が 3.11，2.89，3.67，3.00 と比較的高くなっている．これは，話速が大のときは少し早口にはなっているものの，要領の良さを感じさせる速さになっているが，話速が小のときは話し方がのんびりしているように感じたため，やる気がなく，低信頼感判定を得たと考えられる．自然性，発話速度感に関しても話速が大のときに自然性は 2.78，3.00，3.89，3.00，発話速度感は 4.00，4.00，4.33，4.00 と評価値が高い．特に，発話速度感については図 1 から，話速と今回は実験の対象としていないポーズ長のみが関係しているため，話速が大のときに評価値が高くなり裏付けが取れたと言える．

全体の満足度に関しても話速が大のときに 3.22，3.44，3.89，3.22 と評価値が高くなっている．最も評価が高いのはパターン E の声量小-抑揚大-話速大の 3.89 であり，最も評価が低いのはパターン D の声量大-抑揚小-話速小の 2.67 で，要素感覚のレベルが真逆となっている．声量が大のときより小のときのほうが評価値が高くなるの

は，音源であるスピーカーと被験者との距離が 100cm で実験を行ったため，声量が大のときではうるさく感じられ，実験を実験音声のみが聞こえる防音性のスタジオで行ったことによって，声量小のときにちょうどよく感じたからだと考えられる．パターン E はこの他に，親しみやすさ，信頼感，自然性，発話速度感の項目で最も高い評価を得ており，発話印象 5 つの評価が高いほど全体の満足度も高くなるという結果が得られた．

5　要素感覚による好印象発話モデルの構築

西田ら [2] の論文をもとに，各要素感覚のパターンに対してレベル付けを行う．レベルはそれぞれのパターンに対して $A_i = a_1, a_2, \cdots, a_j, \cdots, a_n$ と表すことができる．レベル付けの例を図 4 に示す．

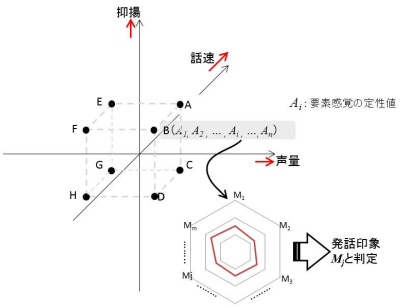

図 4　要素感覚のレベル付け．

これらのレベル付けをもとに，好印象発話モデルの構築方法を検討する．声量が大となるパターンにおいて，熱意の評価値が高くなり，自然性の評価値は低くなった．また，話速が大となっているパターンは発話印象の評価値が高くなっていたため，話速は全ての発話印象に相関があると考えられる．反対に，抑揚に関してはレベルに関わらず評価が一定でなく，特に傾向が見られなかったため，今回の合成音声を用いた実験においては相関が見られないと考える．今回の実験結果から考察した好印象発話モデルを図 5 に示す．要素感覚が大のとき発話印象の評価値が大きくなる場合は実線，小さくなる場合は点線で示している．

6　まとめ

本稿では好印象発話モデルを構築するために，音声評価実験を行った．南らが提案したモデルのうち要素感覚と発話印象についての関係性を示すため，様々な組み合わせの要素感覚を持つ合成音声を聴かせ，発話印象の評価をしてもらうという実験を行い，声量の大小により熱意，自然性の評価値が変化する，抑揚の大小はどの発話印象の評価値にも影響を与えない，話速は全ての発話印象の評価値が高くなる，という結果を得た．その結果をもとに，好印象発話モデルのうち，要素感覚と発話印象の関係例を示した．今後，本実験において用いなかった要素感覚である，明瞭性，ポーズ長を加えた合成音声を作成し実験を行うこと，被験者を増やしデータに更なる信憑性を持たせること，要素感覚を大小の 2 段階だけでなく，より細かく段階をわけることなどが課題として挙げられる．

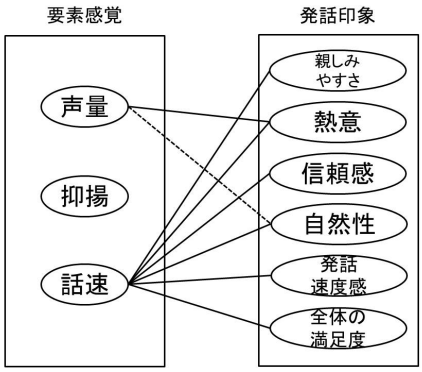

図 5　好印象発話モデル．

参考文献

[1] 南寛輝. 音声情報を用いたプレゼンテーションに対する反応モデルの構築. 長岡技術科学大学修士論文, 1 2014.

[2] 西田悠，浅田龍星，中平勝子，北島宗雄. 発話音声から受ける要素感覚を決める音響特徴量の評価指標.FIT2017(印刷中)，9 2017.

[3] 籠宮隆之・山住賢司・槙洋一・前川喜久雄，講演音声の大局的な印象に影響を与える要因，音声研究，Vol.11，No.2，pp.65-78 (2007).

[4] 内田照久，音声中の抑揚の大きさと変化パターンが話者の性格印象に及ぼす影響，心理学研究，Vol.76，No.4 pp. 382-390 (2005).

[5] 西崎博光・関口芳廣，教員の話し方改善システムの開発に向けた講演音声の特徴分析，日本教育工学会論文誌，Vol.34，No.3 pp.171-179 (2010).

[6] 騒音値の基準と目安，http://www.skklab.com/standard_value

[7] dB の話し　音の大きさ，http://www.geocities.jp/fkmtf928/dB_sound.html

[8] 小宮山将紀. 発話速度とポーズ長が文章理解に与える影響. 千葉工業大学.