音節と継続時間を考慮した擬音語決定手法の提案

A Proposal for an Onomatopoeia Determination Method Considering Syllable and Duration of Sounds

塩澤 朋[1*] 長谷川 大[2] 佐久田 博司[1]
Tomo Shiozawa[1] Dai Hasegawa[2] Hiroshi Sakuta[3]

[1] 青山学院大学大学院 理工学研究科 理工学専攻
[1] Graduate School of Science and Technology, Aoyama Gakuin University
[2] 東京工科大学 メディア学部
[2] School of Media Science, Tokyo University of Technology
[3] 青山学院大学 理工学部
[3] College of Science and Engineering, Aoyama Gakuin University

Abstract: Avatar-Mediated Communication (AMC) において，環境音を擬音語で伝達することで様々な表現が可能になる．本稿では，環境音に対応した擬音語を表示させることで，環境音の可視化を行い，視覚的に音を伝達するための手法を提案する．提案手法では，音を音節に分割し，スペクトル構造を調べることで，擬音語への変換を可能にする．表示される擬音語のふさわしさを評価し，提案手法の有効性を確認する．

1 はじめに

1.1 背景

近年，Avatar-Mediated Comuunication(AMC) と呼ばれる，アバタを介したコミュニケーションが発展している．そういったアバタや仮想空間を介したコミュニケーションにおいて，環境音の表現についてはあまり議論がされておらず，環境音の情報はそのまま音声として伝達されている．そこで，擬音語に注目し，擬音語表現を表示させることで環境音を視覚的に伝達することを考える．また，環境音を擬音語表現を利用して伝達することで AMC における環境音の表現の幅が広がるといえる．例えば，3D アバタを媒介として見守りシステム [1] などのような，プライバシーを配慮する必要がある場合など，音情報をそのまま伝達できない状況においても，環境音を擬音語表現に変換することで，音情報をそのまま伝達するよりもプライバシーに配慮した伝達が可能となる．

擬音語は音を字句で表現したものであり，日常会話などでも音を伝達するために利用されており，環境音を表現する有効な方法であると言える．しかし，AMC で擬音語表現を利用した時には，環境音から擬音語への変換の方法や擬音語表現の表示によって正しく伝達ができるのかといったことを検討しなければならない．そこで我々は環境音に対する擬音語の決定手法に関しての検討を行う．

1.2 関連研究

環境音に対する擬音語の決定に関する研究はこれまでにも様々なものが行われている．石原らは環境音に対する擬音語の自動認識を試みている [2]．石原らの手法では，入力された環境音の波形データを音節単位に分割し，MFCC を用いてその特徴量と擬音語を構成する音素との対応付けをしている．しかし，環境音の波形には音声認識における母音や子音にあたる明確な波形がないため，擬音語に対応した音素を設計する必要がある．また，音節の時間長に関係なく音素の対応付けをしているため，継続時間についての考慮がされていない．

比屋根らは，音の最大周波数とそれらの擬音語表現の対応関係について調査している [3]．この研究では，擬音語表現が環境音のスペクトル構造に応じて変化することを示しており，擬音語表現が環境音のスペクトルや継続時間から決定されることを示している．ただし，比屋根らはあくまで環境音の特徴と擬音語表現の対応について調査をしただけであり，擬音語への変換手法については提示していない．

*連絡先： 青山学院大学大学院理工学専攻
〒 252-5258 神奈川県相模原市中央区淵野辺 5-10-1
E-mail: c5616155@aoyama.jp

さらに我々は，先行研究において環境音に対して単発音表現の擬音語をアニメーション表示させるシステムを開発した [4]．このシステムでは，擬音語を利用することで環境音を可視化し，環境音を視覚的に伝達することを試みた．

しかし，石原らや我々の先行研究では擬音語の決定手法において，環境音の継続時間を考慮していないという課題があった．前述のシステムでは環境音の継続時間を考慮していなかったために,「ドン」や「パン」といった継続時間の短い環境音を表現する擬音語しか表示することができなかった．そのため，取得した環境音の継続時間が長い場合であっても，短い環境音を表現する擬音語が表示されてしまい，残響音の表現としてふさわしくない擬音語が表示されてしまった．そのため，環境音の継続時間を正しく伝達することができていなかったと言える．したがって，残響音などの環境音に対しても継続時間を反映した擬音語表現を表示する必要があり，そのためには擬音語の決定手法の改善が必要である．

1.3 目的

そこで本研究では，環境音の継続時間も考慮した擬音語の決定手法の提案を行い，提案手法を実装した環境音の可視化システムを開発する．先行研究の手法によって表示される擬音語表現と，提案手法によって表示される擬音語表現を比較することで提案手法の有効性を評価する．

2 提案手法

本研究では，先行研究での手法を改善し，環境音の継続時間を考慮した擬音語の決定手法を提案する．提案手法では，1 つの擬音語を語頭表現と語尾表現に分け，環境音の継続時間に合わせて語尾表現を変更することで環境音の長さを表現する．おおまかな提案手法のおおまかな流れは図 1 の通りである．提案手法では，取得した環境音の波形データを音節単位に分割する．次に，分割した音節のスペクトル解析を行い擬音語の語頭表現を決定する．そして音節の継続時間に合わせて語尾表現を決定し，決定された語頭表現と語尾表現を合わせることで 1 つの擬音語表現とする．

2.1 対象とする擬音語

環境音には，うなり音や連続音，単発音など様々なものがあり，それを表現する擬音語表現にも多種多様なものが存在する．

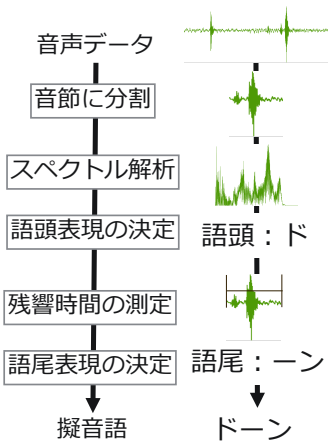

図 1: 提案手法の流れ

提案手法で対象とする擬音語は，単発音と呼ばれる短時間に減衰する環境音を表現した擬音語とする．田中らの分類に基づくと，単発音には減衰音，残響音，2 回衝突音，破砕音があり，本稿では減衰音と残響音を対象とする．[5]．減衰音は音圧がすぐに減衰するもので,「コツ」「カチ」といった擬音語表現で表されることが多い．残響音は残響があるもので,「カン」「キン」といった擬音語表現で表される．2 回衝突音の場合には「ガタン」，破砕音の場合には「バリ」のような擬音語表現が用いられる．

2.2 音節への分割

本手法ではまず，取得した環境音の波形データを音節単位に分割する．環境音の波形データに関しては，マイクで取得された一定時間分の音声データを使用する．音声データについては，今回はサンプリングレートを 22,050[Hz]，サンプル数を 8,192 としており，0.372 秒ごとに取得したデータを音声データとして入力する．また，波形データは最大値 1.0，最小値-1.0 の波形データになるように変換している．音節に分割する際にはまず，波形データから音圧の移動平均を算出し，閾値以上になった場合に音節の開始とする．移動平均の区間数は 100 サンプルとしている．この移動平均の値が 0.05 よりも大きくなった場合に環境音の開始と判断する．次に，移動平均の値が，環境音の開始時の移動平均の値の 8 分の 1 になった場合に音節の終了とし，開始から終了までの波形データを 1 つの音節の波形データとする．移動平均の区間数について今回は 100 サンプルとしているが，今後は実験等で最適な値を検証する必要がある．また，閾値の 0.05 という値は，デシベル換算をすると 25[dB] ほどの値である．

表 1: 最大周波数と語頭表現の対応表

最大周波数	語頭表現
150Hz 未満	ド
150Hz 以上　200Hz 未満	ボ
200Hz 以上　600Hz 未満	ト
600Hz 以上 1,000Hz 未満	ポ
1,000Hz 以上 1,500Hz 未満	コ
1,500Hz 以上 2,500Hz 未満	パ
2,500Hz 以上 3,500Hz 未満	カ
3,500Hz 以上 4,500Hz 未満	キ
4,500Hz 以上 5,500Hz 未満	ピ
5,500Hz 以上	チ

2.3　音節のスペクトル解析

環境音を音節単位に分割した後，音節のスペクトル解析を行い語頭表現を決定する．まず，それぞれ音節の波形データに対して高速フーリエ変換（以下，FFT）を行い，周波数スペクトルに変換する．FFT を行う際には，音節の波形データに窓関数としてハミング窓を掛け合わせている．また，ハミング窓の長さは音節の長さに応じて変化させている．そして，得られた周波数スペクトルの中で最大音圧である周波数（以下，最大周波数）に応じて語頭表現を決定する．語頭表現の決定では，10 種類の語頭表現の中から最大周波数に対応する語頭表現が選ばれる．10 種類の語頭表現と最大周波数の対応は表 1 のようになっている．比屋根らは単発音の最大周波数と擬音語の関係を調査し、10 種類の語頭表現が対応する傾向にあることを確認しており，それを参考に提案手法では 10 種類の語頭表現としている [3]．また，語頭表現に用いる表現の選定および最大周波数と擬音語表現の対応については，比屋根らの研究に述べられている周波数と擬音語の関係を参考にしている [3]．

2.4　音節の継続時間

次に，音節の継続時間に応じて語尾表現を決定する．音節の波形データの開始から終了までの時間を環境音の継続時間として，継続時間の長さに対応した語尾表現を決定する．語尾表現には 3 種類のパターンがあり，それぞれ促音（ッ），發音（ン），長音素と發音（ーン）となっている [3]．継続時間と対応する語尾表現は，表 2 のようになっている．また，継続時間と語尾表現の対応は予備実験の結果を元にしている．予備実験は，継続時間だけを変更した音を実験参加者に聴取させ，3 パターンの語尾表現の中からふさわしいと思う語尾表現を選択してもらうものである．聴取する音は，周波数が 180[Hz] の正弦波で，継続時間を 10[ms] から 400[ms] まで 10[ms] ごとに変化させている．予備実験の参加者は 20 代の男性 7 名，女性 1 名である．予備実験の結果を表 3 に示す．表 3 はそれぞれの継続時間において，どのパターンの語尾表現を何人の参加者が選択したかをまとめたものである．継続時間について，比屋根らは周波数 4,000[Hz] での音について調査を行っており，本研究の予備実験の結果と差異が見られる [3]．これは周波数の違いによるものと考えられ，今後も周波数の違いによる語尾表現の変化などについて検討していく必要がある．

表 2: 残響時間と語尾表現の対応

残響時間 [ms]	語尾表現
0〜170	ッ
170〜270	ン
270〜	ーン

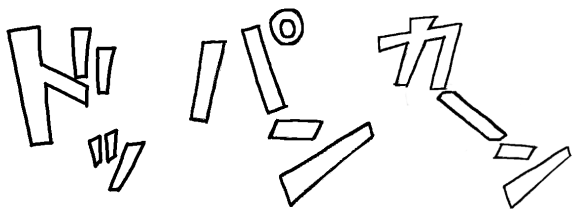

図 2: 表示される擬音語の例

2.5　擬音語の決定

最後に，決定した語頭表現と語尾表現から 1 つの擬音語を決定する．語頭表現の計 10 種類と語尾表現の計 3 種類の組み合わせにより，30 種類の擬音語表現のパターンから 1 つの擬音語が決定される．後述のシステムで表示される擬音語の例を図 2 に示す．

3　システム概要

本研究で開発したシステムの構成及び処理手順を図 3 に示す．システムでは，Kinect for Windows V2 を使用して環境音を取得する．そして，取得した環境音に対して提案手法を用いて対応する擬音語表現を決定し，カメラの実映像に重ねて擬音語表現を表示させる．また，擬音語は環境音の方向に合わせて表示させ，音の大きさに合わせて表示する擬環境音語のフォントの大きさも変化するようになっている．したがって，本

表 3: 継続時間と語尾表現の対応についての予備実験結果

時間 [ms]	ッ	ン	ーン	時間 [ms]	ッ	ン	ーン	時間 [ms]	ッ	ン	ーン	時間 [ms]	ッ	ン	ーン
10	8	0	0	110	6	2	0	210	5	3	0	310	0	1	7
20	8	0	0	120	7	1	0	220	1	5	2	320	0	1	7
30	8	0	0	130	6	2	0	230	1	2	5	330	0	1	7
40	8	0	0	140	7	1	0	240	0	5	3	340	0	0	8
50	8	0	0	150	5	3	0	250	0	5	3	350	0	2	6
60	8	0	0	160	6	2	0	260	0	4	4	360	0	1	7
70	7	1	0	170	6	2	0	270	0	5	3	370	0	0	8
80	7	1	0	180	3	5	0	280	1	2	5	380	0	0	8
90	7	1	0	190	1	6	1	290	0	2	6	390	0	1	7
100	7	1	0	200	4	4	0	300	0	1	7	400	0	1	7

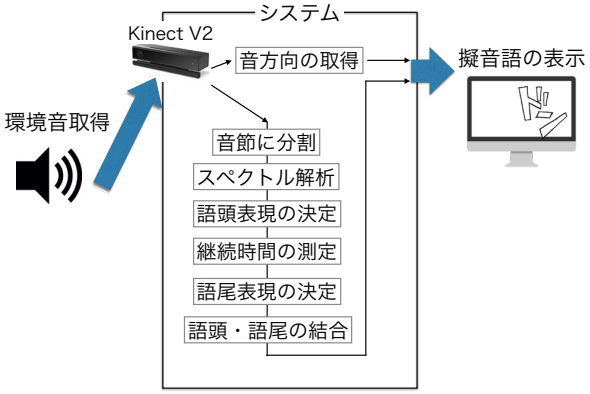

図 3: システム構成及び処理手順

システムは取得した環境音に対して，どのような音であるか，音量，音源方向を視覚的に伝達できるものとなっていると言える．システムの実行画面を図 4 に示す．ただし本稿では，音の方向や擬音語表現のフォントや大きさなどの要素は検討せず，今後検討していく予定である．

4 実験

提案手法でのシステムの方が先行研究のものよりもふさわしい擬音語表現を表示できているかどうかを評価するために実験を行った．実験では，実験参加者にシステムの実行画面を録画したビデオを見てもらう．参加者は，実際の環境音とシステムで表示された擬音語表現を比較して，システムが実際の環境音にふさわしい擬音語表現を表示できているかを評価する．

評価はアンケート方式で集計し，1 つのビデオに対して語頭表現と語尾表現のそれぞれについてふさわしいかどうかを 1 から 7 の 7 段階評価となっている．

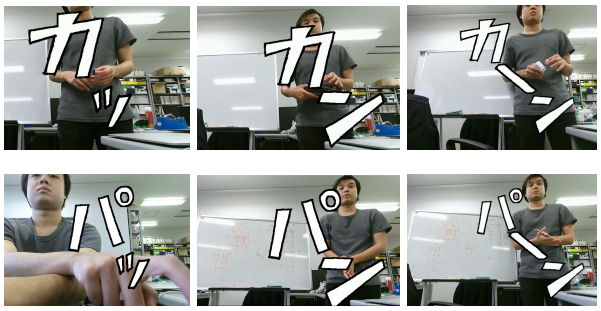

図 4: システムの実行画面

ビデオは，1 つのビデオにつき 1 種類の環境音が録画されており，手を叩く音，呼び鈴の音など 8 種類である．また，提案手法と先行研究での評価を比較するために，1 種類の音に対して提案手法のものと先行研究のものでビデオを作成した．ビデオの提示順は，環境音の種類と手法の違いに関わらず全てランダムで提示した．さらに，参加者には手法がどちらであるのかを分からないように提示した．実験の参加者は，20 代の男女 19 名である．

5 結果

実験の結果を表 4 と表 5 に示す．表 4 は各手法の語頭表現とその評価をそれぞれの音と全体平均でまとめたものである．表 5 は同じく語尾表現の評価をまとめたものである．表では，先行研究での手法を旧手法としている．

結果として，全体でみると語頭表現の評価が提案手法の方が高い評価となっており，有意差がみられた．語尾表現の方には有意差が出ず，提案手法と旧手法の全体での評価は変わらなかった．

それぞれの環境音について，語頭表現に関しては，

表 4: 語頭表現の平均評価

環境音の概要	提案手法		旧手法	
手を叩く	パ	6.74	パ	6.74
プラスチック箱	ポ	5.00	ポ	5.26
ページめくり	パ	2.53	カ	1.79
ホチキス	パ・チ	4.26	パ	2.95
呼び鈴	チ	5.47	カ	3.32
硬貨を落とす	パ・コ	3.21	コ	3.68
机を叩く	ト	6.11	ト	6.32
テニスボールのキャッチ	コ	4.00	コ	4.63
全体平均	-	4.66	-	4.33

表 5: 語尾表現の平均評価

環境音の概要	提案手法		旧手法	
手を叩く	ッ・ン	6.11	ン	6.16
プラスチック箱	ッ	4.74	ン	6.11
ページめくり	ッ	3.47	ン	2.26
ホチキス	ッ・ン	4.11	ン	3.53
呼び鈴	ーン	6.26	ン	5.16
硬貨を落とす	ーン	3.68	ン	4.21
机を叩く	ッ	5.21	ン	6.53
テニスボールのキャッチ	ッ	5.74	ン	5.89
全体平均	-	4.97	-	4.98

ページめくり，ホチキス，呼び鈴の 3 種類の環境音で提案手法の方が高い評価を得ており，有意差がみられた．一方で，テニスボールをキャッチする音については有意差はみられないものの，提案手法の方が旧手法よりも評価が 0.63 ほど低くなっている．語尾表現に関しては，語頭表現と同じく，ページめくり，ホチキス，呼び鈴の 3 種類で提案手法の方が高い評価であり有意差もみられた．逆に，プラスチック箱を叩く，机を叩く音の 2 種類では旧手法の方が評価が高くなっており，提案手法の評価が優位に低くなっている．

6 考察

語頭表現の評価について，最大周波数と語頭表現の対応に関しては提案手法と旧手法で違いが無いにも関わらず有意差がみられた．これは，旧手法ではスペクトル解析の前に音節への分割を行っていないが，提案手法では音節への分割を行っているため，語頭表現に違いが生じたためだと考えられる．スペクトル解析の前に音節への分割を行ったことで，結果的に余計な波形データが取り除かれることになり，ノイズの少ないより正確な分析が可能になったと考えられる．そのため旧手法よりも提案手法の方が，よりふさわしい語頭表現を選択することができたと言える．

語尾表現については，全体でみると差は見られなかった．これは長音素のパターンは評価が高いのに対し，促音のパターンでは評価が低くなってしまったためである．促音のパターンの評価が低い原因は，音の継続時間の決定手法に問題があること，音の継続時間と語尾表現の対応が不適切であることが考えられる．音の継続時間の決定手法については，提案手法では音節の長さを音の継続時間としているため，音節へ分割する部分で課題があると考えられる．そのため，提案手法での音節への分割する部分を見直して，音節が適切に区切れているかどうかを評価する必要があると言える．継続時間と語尾表現の対応については，促音パターンの語尾表現の評価が低くなっているため，促音パターンと撥音パターンの区切りを再検討する必要がある．促音パターンよりも旧手法での撥音パターンの評価が高くなっていることから，促音パターンとなる継続時間をより短く設定することで改善することができると考える．また，継続時間と語尾表現の対応関係を再度調査する必要がある．今回は簡単な予備実験での調査であったため，より厳密に調査を行えばより評価の高い対応関係が得られるかもしれない．

今後は，考察にあるような改善を行っていき，より高評価を得られる決定手法を目指していく．

参考文献

[1] 長谷川大, 小林裕, 白川真一, 佐久田博司, 安彦智史, 安達栄治郎, 中山栄純: アバタ媒介型見守りシステムの開発, 知能と情報 *(*日本知能情報ファジィ学会*)*, Vol.28, No.6, pp.974-85 (2016)

[2] 石原一志, 駒谷和範, 尾形哲也, 奥乃博: 環境音を対象とした擬音語自動認識 擬音語表現における音素決定あい昧性の解消, 人工知能学会論文誌, Vol.20, pp.229-236 (2005)

[3] 比屋根一雄, 澤部直太, 飯尾淳: 単発音のスペクトル構造とその擬音語表現に関する検討, 電子情報通信学会技術研究報告. *SP,* 音声 , Vol.97, No.586, pp.65-72 (1998)

[4] 塩澤　朋, 長谷川　大, 佐久田　博司: 擬音語を利用した室内環境認知支援システムの試作, インタラクション *2016* 論文集, pp.814-818 (2016)

[5] 田中基八郎, 松原謙一郎, 佐藤太一: 異音の表現における擬音語の検討: 衝突音等の単発音やうなり音

の場合, 日本機械学会論文集 C 編, Vol.61, No.592,
pp.4730-4735 (1995)