

オノマトペ発話の変化に着目したラインアニメーション表現

Line Animation Focused on Onomatopoeia Pronunciation Changes

甲田春樹*1 佐藤宏介*1
KODA Haruki SATO Kousuke

*1大阪大学 大学院基礎工学研究科
Graduate School of Engineering Science Osaka University

When uttering onomatopoeia, a speaker varies volume and pitch intentionally and unintentionally. However, it is difficult to associate speech characteristics such as volume or pitch and speakers intention, because these characteristics depend on the attributes of the speaker and recording conditions. Through an experiment where subjects speech onomatopoeia synchronized with presented line animations, we investigated the correspondence between changes in the animation design parameters and changes in verbal and non-verbal information.

1. はじめに

計算機器の演算処理の向上や音声認識技術の発展を受け、携帯端末機やカーナビゲーションシステムでの音声入力機能が実用化されて久しい。音声入力ではコマンド選択やテキスト入力が主目的な為、話者性や音量・音程等の非言語情報は、抽出すべき言語情報から分離・除去される事が望ましいとされている。しかし、会話において人間は発話の音量・音程を豊かに変化させ、その話し方は話者毎に特有であり、このような非言語情報に話者の表現意図が含まれている事は想像に難くない。非言語情報の変化を捉え理解する事で、認知的負担の少ない自然な音声入力インタフェース設計の知見を得られる可能性がある。

そこで我々は、擬音語や擬態語であるオノマトペが、一般単語に比べて発話時の非言語情報の変化に富んでいる点に着目した。オノマトペは物の動きや様態を記述する言葉であり、単語内に長音や繰り返しの構造を含む物が多い。前者の例としては「びゅーん」や「ふわーん」、後者の例としては「グルグル」や「ピカピカ」などが挙げられる。単語の担う意味が感性的であり、特徴的な構造を持つ事が、オノマトペの発話継続時間の増減や音量音程の変化に富む理由として考えられる。

しかし、非言語情報は個人差や個人内の発話毎の変動が大きく、また収録環境の再現性の低さもインタフェースを設計するにあたり課題の一つである。我々は、非言語情報と表現意図を一義的に対応付けするのではなく、発話間の相対的な変化から意図を読み取れるのではないかと考えている。

本研究では、発話による表現対象として、線画アニメーションの各種デザインパラメータとの対応付けを恣意的に仮定した上で、非言語情報の相対的な変化を観察する実験を行った。線画とオノマトペの相性の良さは経験的に知られており、幼児が描画作業の際にオノマトペを発話する事、漫画の効果線表現にしばしばオノマトペが伴う事が例として挙げられる。具体的な手続きとしては、線画アニメーションを繰り返し提示した時の発話の変化を観察することで、音声のインタフェース入力としての適性および、音響的特徴量の変化とデザインパラメータの変化の間の対応関係についての知見を得た。また、知見を基にして音声でドローイングのデザインパラメータを対話的に操作するアニメーション作成環境構築を試作した。

2. 関連研究

オノマトペや音声入力インタフェースに関する関連研究を、本稿では以下の2種に分類して述べる。

2.1 音声の非言語情報を利用したもの

非言語情報をインタフェースに利用した例としてまず、藤井らのコエカキ [Fujii 05] を挙げる。コエカキは鑑賞者の声に対応してドローイングが描かれるメディアアートであり、音程や音量の変化に同期してドローイングの色合いや太さ、方向等の各種パラメータが変化する。次に、五十嵐らは音声による直接操作インタフェース [Igarashi 01] を提案しており、発声の有無によるオン・オフ操作あるいはタンギングによる離散的パラメータ調整を実現している。これらの研究では、ハンズフリーかつ直感的なインタフェースを実現する事が目的であり、入出力間の対応付けは経験的に設定されている。

2.2 オノマトペの言語情報を利用したもの

オノマトペが保持する意味や意図を明らかにしようとする研究としては、言語情報から一義的に対応づけられる音象徴と動作やデザインパラメータとの対応付けを行なうものが幾つか報告されている。

木島らは、オノマトペ-動作特性間の共時的対応を物理変数で確認する実験を行った [Kijima 06]。具体的には、/g/u/や/k/j/u/などのオノマトペを音声で提示した被験者に把持動作を行わせ、音韻構造と動作の出力特性の対応関係を確認した。動作との対応付けを行なった他例として、小松らの研究 [Komatsu 09] が挙げられる。小松らはまず、子音や母音等の音節の構成要素に対して属性ベクトルを設定し、ベクトルに基づいて単語全体の印象を表現した。その後、各単語の属性ベクトルとロボットのモーションを対応づける事で、オノマトペによりロボットのモーションを指示するシステムを構築した。また、神原らのオノマトペ [Kambara 10] では、ユーザがドローイングを行いながら「キラキラ」「もこもこ」といった発話を行う事で、ブラシの質感を変化させる事が可能である。

これらの先行研究は、言語情報(およびそれに一義的に対応づけられる音象徴)が担う感性情報を利用して動作や質感との対応関係の明らかにしようとしているものである。人間は日常会話にオノマトペを用いて意思・意図伝達を行えている事からも、各単語になんらかの普遍的な感性情報が存在する事は妥当だと言える。それに対し我々は、個々の単語の持つ意味や情報

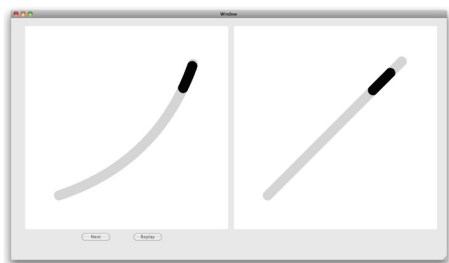


図 1: 実験用アプリケーション 2つのウィンドウ (600 × 600[pixel]) および next・replay ボタンで構成される

ではなく、前後して発話されたオノマトペの非言語情報の変化から話者の表現意図を読み取れるのではないかと考えている。

3. 線画アニメーション提示による発話実験

音声のインタフェース入力としての適性および、オノマトペ発話時の発話特性の変化を調べる為の実験を実施した。実験では、被験者に線画アニメーションを提示し、それに伴う発話を指示した。実験1においては/a/音、実験2においては「びゅーん」「ふわーん」の2種の擬音語を用いるよう指示した。

3.1 実験の環境と手続き

実験環境として、刺激提示用コンピュータ (Apple社 MacBookPro, Intel Core i5, 2.4GHz, 4GB RAM, 110[ppti]), およびヘッドセットを用いた実験用アプリケーション (図1) を構築し、音声については表1に示す条件で録音・分析を行った。有声区間の検出は閾値法により行い、閾値は録音データの目視により決定した。アプリケーションはアニメーションを表示する二つのウィンドウおよび「next・replay ボタン」から構成され、被験者自らが操作を行うよう指示した。各試行は基準刺激と比較刺激の組からなり、被験者はまず next ボタンを押して表示される基準刺激に対応する発話を行い、その後、再度ボタンを押して表示される比較刺激に対応する発話を行う。刺激に対し再度発話を行う場合には replay ボタンを押す。

3.2 提示刺激パターン

提示刺激として、速度・線幅・線長・形状をデザインパラメータとする線画アニメーションを作成した。アニメーションの概観および各パラメータの詳細をそれぞれ、図2、表2に示す。形状は2次ベジェ曲線により作成し、始点・終点は固定としたまま制御点の座標値を変える事で5種の形状を用意した。基準刺激は [速度, 線幅, 線長, 形状]=[中速, 中, 中, 直線] とした。比較刺激については、各実験の説明にて後述する。

3.3 実験1: 非言語情報とドロージングの対応

実験1においては、/a/音による発話を指示し、提示刺激として [速度, 線幅, 線長, 形状] のうちパラメータの一つを変化させたパターン16種を比較刺激として用いた。よって、本

表 1: 音響分析条件

サンプリング条件	41.1[kHz] / 32[bit]
フレーム窓	Hamming window
フレーム長	2048/44100 [sec] (約 46[ms])
シフト長	512/44100 [sec]
音量	音圧 512 点の絶対値の平均 [dB SPL]
基本周波数の抽出	ケプストラム法
フォルマント (f1, f2)	分析ソフト WaveSurfer を使用

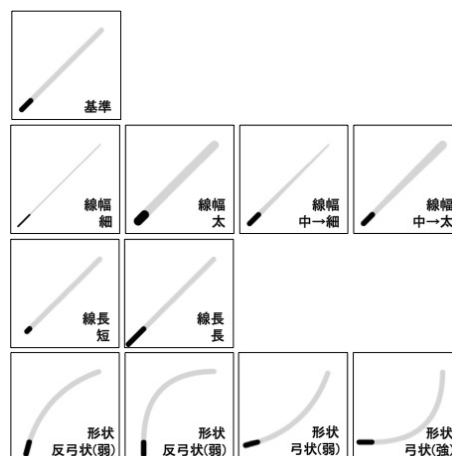


図 2: 提示刺激 (抜粋) 速度および線長が時間変化する刺激の掲載については割愛とした

実験では基準刺激と比較刺激の組み合わせ16組からなる。被験者は20代の男女4名である (男性: 2名, 女性: 2名)。

3.4 実験1の結果と考察

3.4.1 パラメータ入力法としての音声の適性

各被験者が基準刺激に対して発話した全16試行に対し、発話開始の遅延および発話継続時間それぞれの平均と標準偏差を表3に示す。本実験では開始されたアニメーションに合わせて発話を開始するため、平均で0.47[s]の遅延が見られた。デザインパラメータの入力として音声を利用する場合、入出力間で多少の時間差が許される作業環境ならば、遅延の補償機能をアプリケーションに実装する事は可能であると考えられる。一方、各被験者の遅延の標準偏差は平均で0.12[s]であり、これは30[fps]のアニメーションにおいては3,4フレームに相当する。今回の実験は少数の被験者が対象であるが、この傾向が多数に対しても当てはまる場合、数フレーム毎に調整が必要なパラメータの入力においては、音声の発話開始遅延の偏差は操作性に影響を与えないと考えられる。

一方発話継続時間については、基準刺激の再生時間が1.88[s]であるのに対し、各被験者の平均は1.85[s]、標準偏差は0.16[s]となった。平均して0.47[s]の遅延が生じたにも関わらず、再生時間と発話継続時間がほぼ一致している事から、刺激終了後も発話が継続して行なわれていた事が分かる。習字の「払い」を表現する様な、動作終了前後に細かいパラメータ操作が求められるラインアニメーションインタフェースを設計する場合、このような発話の継続効果を利用したパラメータ操作法を実現できる可能性がある。

3.4.2 基準刺激・比較刺激提示時の音量・音程の相対変化

各被験者が基準・比較刺激それぞれに対して行なった発話における、有声区間時の音量・音程の散布図を図3に示す。特徴

表 2: 線画アニメーションのデザインパラメータ 太字は基準刺激に用いたパラメータを示す

	速度	低速	中速	高速	減速	加速
[pixel/sec]		150	300	500	300 → 150	300 → 500
線幅		細	中	太	中 → 細	中 → 太
[pixel]		5	15	25	15 → 5	15 → 25
線長		短	中	長	中 → 短	中 → 長
[pixel]		50	100	150	100 → 50	100 → 150
形状		直線	弓状 (弱)	弓状 (強)	反弓状 (弱)	反弓状 (強)

表 3: 発話開始遅延および発話継続時間

	M-1	M-2	W-1	W-2	ALL
発話開始遅延平均 [sec]	0.36	0.44	0.53	0.57	0.47
発話開始遅延標準偏差 [sec]	0.14	0.07	0.12	0.16	0.12
発話継続時間平均 [sec]	2.09	1.74	1.77	1.80	1.85
発話継続時間標準偏差 [sec]	0.18	0.16	0.16	0.33	0.16

量は区間内にて 50[samples/s] でサンプリングをした。比較刺激と基準刺激とでは有声区間時の総サンプル数が異なるが、最もサンプル数に差がある被験者で、比較刺激に対する基準刺激のサンプル数の割合は 0.86 であった。

まず、基準刺激に対する各被験者の傾向を見る。M-2,W-1,W-2 の 3 名は、基準刺激時に一定音量・音程に保とうとする発話が見られたため、分布においても比較的類似していると言える(女性被験者においてはピッチの抽出誤りにより、100[Hz] 周辺にも分布が広がっている)。一方 M-1 はアニメーションが終端に近づくほど音程・音量が増加する発話が見られたため、分布が大きく広がる結果となった。

次に比較刺激時の傾向としては、音程の変化が乏しかった W-1 をのぞき、やや右上に傾いた縞状の分布が確認できる。各試行毎の特徴量は個々の縞にある程度まとまっていると考えられる。この分布は以下の 2 点を示唆すると我々は考えている。1 点目は、比較刺激時にはその刺激種を問わず音量・音程が増加傾向に有る事。そして 2 点目は、音量と音程に相関が見られる事である。基準刺激では音量方向に広がるので分布が見られるのに対し、基準刺激・比較刺激ともに音程方向へ広がる分布が見られない事からも、話者が音程の増加を意図した場合は音量の増加も伴うのではないかと考えられる。

3.4.3 デザインパラメータ変化時の音量・音程の相対変化

線長・線幅が時間変化する刺激を提示した際の各被験者の音量・音程の散布図を図 4,5 に示す。線幅変化刺激は増加・減少刺激ともに線幅が 15[pixel] から開始され、それぞれ単調増加・減少して線幅が 25・5[pixel] になった時点で提示が終了する。線長変化刺激においても同様である。各刺激ともに提示時間は 1.88[sec] である。(刺激提示直後はまだ発話開始されていないため、グラフ上の刺激開始周辺部には分布が存在していない)

全被験者において線幅増加時には音量・音程が、線長増加時には音程が増加傾向が見られる一方、他の刺激では共通する傾向は見られなかった。試行後に被験者らに質問を行なった所、線幅が単調減少する刺激に対して「車が遠方へ高速移動している印象」や「だんだん力がなくなっていく印象」といった様に感性的な判断が異なっていた。これは、音声特徴量変化の傾向が一致しない要因の一つとして考えられる。

3.5 実験 2 : 言語情報とドロ잉の対応

実験 2 においては、オノマトベ「びゅーん」および「ふわーん」による発話を指示した。比較刺激として速度と形状のパラメータのみを変化させたパターン全 14 種を比較刺激として用いた。形状のパラメータには表 2 の 5 種を用い、速度のパラメータには、中速、加速、減速の 3 種を用いた。特徴量として有声区間内のフォルマント (f1,f2) を 100[samples/s] 間隔で記録した。被験者は実験 1 の参加者と同一の 20 代男女 4 名である。

3.5.1 基準刺激・比較刺激提示時のフォルマントの相対変化

各被験者が基準・比較刺激に対して行なった発話での、有声区間時の f1,f2 の散布図を図 6 に示す。比較刺激と基準刺激とでは有声区間時の総サンプル数が異なるが、最もサンプル数に差がある被験者で、比較刺激に対する基準刺激のサンプル数の割合は 0.90 であった。

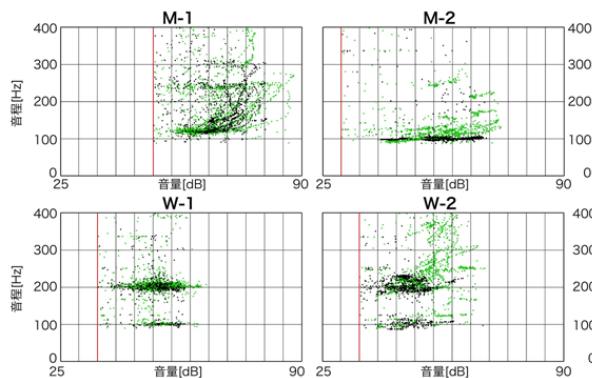


図 3: 基準刺激・比較刺激時の個人毎の音量・音程の分布 黒(緑)のプロットは基準(比較)刺激時の値を示し、赤の縦線は有声区間検出に使用した閾値を示す

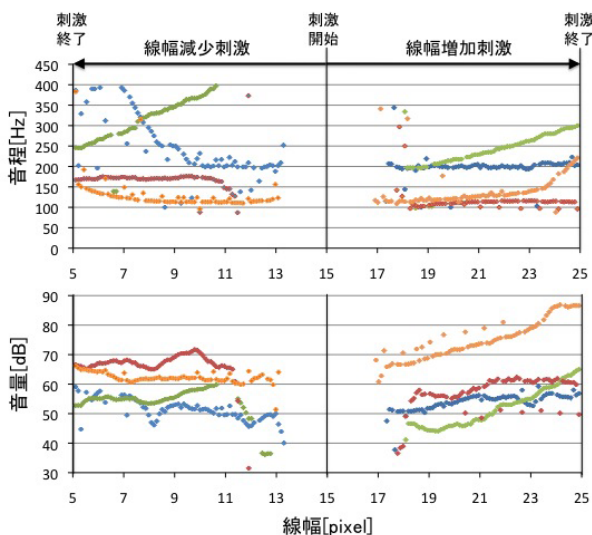


図 4: 線幅変化時の音量/音程の分布 線色は各被験者を示す

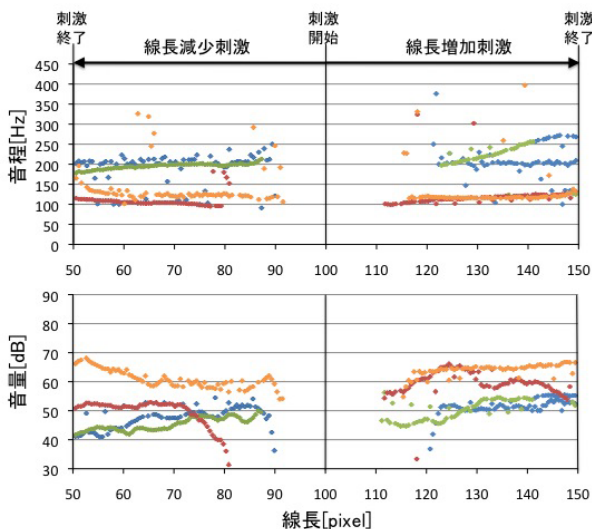


図 5: 線長変化時の音量/音程の分布 線色は各被験者を示す

フォルマントは話者性に大きく左右されるが、各被験者において比較刺激時のフォルマントの分布が基準刺激に比べて

広がっていると言える。また、M-2の「ふわーん」においては比較刺激時のf2の分布が全体に高周波域にシフトしている事や、W-2の「ふわーん」においては比較刺激時のみにf1の300~600[Hz]域に分布が見られる事が分かる。今回の実験では被験者数・試行数ともに限られており、フォルマントの変化とデザインパラメータの対応付けを明らかにするまでは至っていない。しかし、従来の音声認識等において言語情報の特定に用いられてきたフォルマントが、発話者の表現意図を担う情報を保持している可能性を示唆できたとと言える。

4. アプリケーション試作

4.1 音量・音程により線幅を調整するドロワーイングアプリケーション

実験1にて、線幅増加刺激の際、音量・音程が増加する傾向が全被験者に見られた為、音量・音程により線幅を調整するドロワーイングアプリケーションを試作し動作を確認した(図7)。特徴量は30[sample/s]で取得し、入出力のゲイン比は、アプリケーション上のスライダーで調整できる様にした。

音程入力の場合、安定した値の取得が難しい事から瞬時的に線幅が変化するドロワーイングがしばしば観察された。また音量入力の場合、入力値が常に揺らぎを持つため、描画線に独特な線幅変化が表出された。

音声をパラメータ入力に利用するインタフェースの場合、入力値を平滑化する必要があると思われる一方で、音声特有の揺らぎを利用したインタフェースも考えられる。これらについては、実際に作成したアプリケーションの操作性に関するユーザビリティ評価、およびアプリケーションによる作成物への感性的な印象評価の双方の観点から取り組んで行く必要が有る。

5. まとめと今後の展望

音声による線画アニメーション作成環境の構築を想定し2種の基礎実験を行なった。発話の再現性の観点からインタフェース入力として音声の適性を評価し、音量や音程によるデザインパラメータ操作の可能性を示唆した。また、提示刺激が変化する際、発話においてはそのフォルマント分布の分散拡大やシフトが観察された事から、これらの変化とデザインパラメータの変化との対応付けがなされた場合、フォルマントの相対変化をインタフェース入力として利用できる可能性が見い出された。

参考文献

- [Komatsu 09] 小松 孝徳, 秋山 広美: ユーザの直感的表現を支援するオノマトベ表現システム, 電子情報通信学会論文誌 A Vol. J92-A (2009), No. 11 pp.752-763
- [Igarashi 01] 五十嵐 健夫, John F. Hughes: 言語情報を用いない音声による直接操作インタフェース, インタラクティブシステムとソフトウェアに関するワークショップ IX 予稿集 (2001)
- [Kijima 06] 木島 章文, 調枝 孝治: オノマトベの音韻要素が拘束する把持動作の量的・質的パラメータ, 体育学研究 Vol. 51 (2006), No. 5 pp.663-675
- [Kambara 10] 神原啓介, 塚田浩二: オノマトペン, ComputerSoftware Vol. 27 (2010), No. 1
- [Fujii 05] 藤井 伯文: koekaki(コエカキ) 一声で絵を描くー, インタラクション 2005 予稿集,(2005)

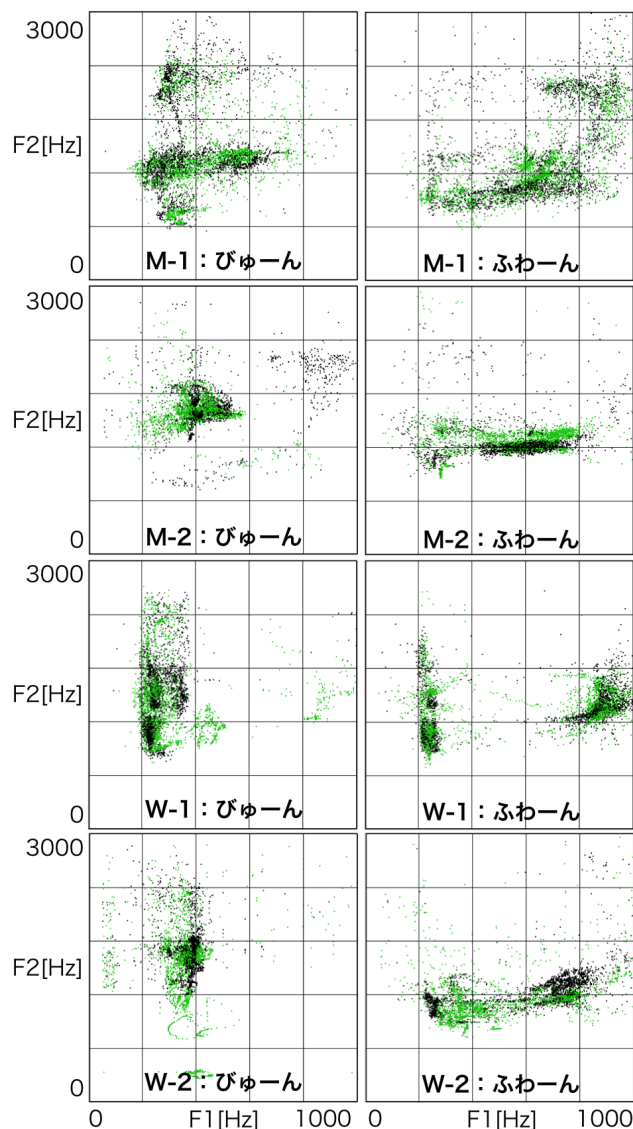


図6: 基準刺激/比較刺激時の個人毎のf1/f2の分布 黒(緑)のプロットは基準(比較)刺激時の値を示す

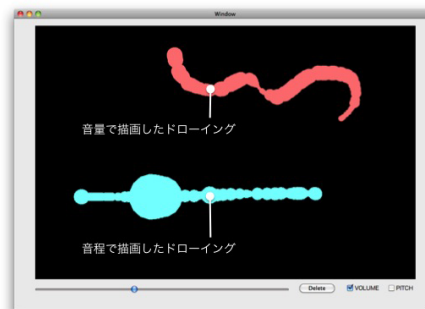


図7: 試作したアプリケーション例 音量・音程により線幅の調整を行なう