

# 調停要約のための正解コーパスの作成とその分析

渋谷 英潔† 中野 正寛‡ 宮崎 林太郎‡ 石下 円香† 永井 隆広‡ 森 辰則†

† 横浜国立大学 大学院 環境情報研究院 ‡ 横浜国立大学 大学院 環境情報学部

E-mail: {shib,nakano,rintaro,ishioroshi,nagadon,mori}@forest.eis.ynu.ac.jp

## 1 はじめに

Web 上には出所が不確かな情報や利用者に不利益をもたらす情報などが存在するため、信頼できる情報を利用者が容易に得るための技術に対する要望が高まっている。しかしながら、情報の内容の真偽や正確性を自動的に検証することは困難であるため、我々は、情報の信憑性は利用者が最終的に判断すべきであると考え、利用者の信憑性判断を支援する技術の実現に向けて調停要約と定義した新しい要約概念に関する研究を行っている [1, 2, 3]。

調停要約とは、一見互いに対立しているようにみえる二言明<sup>1</sup>が、ある状況下で両立可能となるような場合、その状況と両立可能であることを簡潔な文章で提示することである。例えば、利用者が「ディーゼル車は環境に良い」という言明の信憑性を判断するために、この言明をクエリとして一般的な Web 検索エンジンにより検索したとする。この場合、おそらく「ディーゼル車は環境に良い」という主張の文書と「ディーゼル車は環境に悪い」という主張の文書の両方が検索されることになるだろう。一見すると互いに矛盾した内容の文書であるように思われるが、両者を詳細に読んだ結果、一方が「ディーゼル車は二酸化炭素の排出量が少ないので環境に良い」という内容であり、もう一方が「ディーゼル車は粒子状物質や窒素酸化物の排出量が多いので環境に悪い」という内容であったとすると両者の主張は論理的に矛盾しない。このような場合に、検索された文書集合を要約して、最初から「ディーゼル車は二酸化炭素の排出量が少ないので地球温暖化の面では環境に良いが、粒子状物質や窒素酸化物の排出量が多いので大気汚染の面では環境に悪い。環境に良いか悪いかは想定している環境の種類による」といった第三者視点からの簡潔な文章を提示できるシステムがあれば、「環境の種類を明確にしない限り単純に真偽を判断できない」ということを利用者に気付かせることができ、利用者の信憑性判断を支援する上で有効であろう。以上が、我々の調停要約が目指すところである。

調停要約は抜粋型の複数文書要約の一種であるが、要約対象となる文書間で対立しているように見える内容が記述されているという前提の下、その対立の読み解き方を示すことを目的としている点で、一般的な目的の複数文書要約とは異なる特徴をもつ。それゆえ、調停要約を分析・評価するための情報が付与されたコーパス（調停要約コーパス）を入手により作成し、そこで得られた知見に基づいて調停要約自動生成システム HERMeS の研究開発を進めている。本稿では、調停要約コーパスの作成作業を説明した後、調停要約コー

表 1: 調停要約作成作業の流れ

- |                |
|----------------|
| T1. 背景知識の獲得    |
| T2. 対立言明の抽出    |
| T3. 調停知識の獲得    |
| T4. 調停文書の収集    |
| T5. 調停記述の抽出    |
| T6. 情報発信者の抽出   |
| T7. 同一情報発信者の判断 |
| T8. 自由記述要約の作成  |
| T9. 直接調停要約の分類  |
| T10. 直接調停要約の評価 |

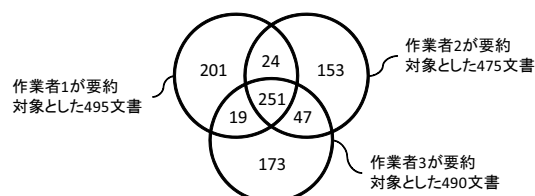


図 1: 要約対象文書集合の重複度

パスを分析して得られた知見に関して報告する。

## 2 関連研究

複数文書間の意味的関係を扱ったコーパスに関する研究には以下のものがある。衛藤ら [4] は、Radev [5] の CST (Cross-document Structure Theory) に基づいた意味的関係を新聞記事集合に付与した文書横断文間関係コーパスを構築した。また、村上ら [6] は、利用者が着目している言明に関連する情報を 6 種類のカテゴリに分類して俯瞰的に提示する言論マップ生成課題 [7] に向けて、Web 上の多様な言明間に意味的関係を付与した言明間意味的関係コーパスを構築した。しかしながら、これらの研究では、文間または言明間同意や対立といった意味的関係を付与することが主な目的であり、その関係の読み解き方に関する記述が文書のどこに書かれているかといった情報は付与されていない。調停要約を自動生成するシステムを開発するにあたっては、そういった読み解き方に関する情報が重要であるため、我々は独自に調停要約コーパスを構築する必要があった。

## 3 調停要約コーパス作成作業

1 章で述べたように、我々は調停要約を情報信憑性判断支援技術の一つとして位置付けており、利用者が信憑性を判断したい言明（着目言明）のみを入力することを想定している。したがって、着目言明と対立す

<sup>1</sup>本稿では、主観的な意見や評価だけでなく、疑問の表明や客観的事実の記述を含めたテキスト情報を広く言明と定義する。

表 2: T3. で選択された対立軸

共通	(a) 飲酒は健康に良い 飲酒は健康に悪い (b) 飲酒は少量ならば健康に良い 飲酒は少量でも健康に良くない (c) 飲酒は心臓病のリスクを上げる 飲酒は心臓病のリスクを上げない	
作業 1	(d) 飲酒は糖尿病に関係ある 飲酒は糖尿病に関係ない (e) 飲酒は快適な睡眠ができる 飲酒は快適な睡眠ができない	他 20 対
作業 2	(d) お酒は脳に良い お酒は脳に悪い (e) 飲酒は糖尿病につながる 飲酒は糖尿病につながらない	他 14 対
作業 3	(d) 飲酒はボケ防止に効果がある 飲酒はボケ防止に効果はない (e) 飲酒は寿命を延長する 飲酒は寿命を延長しない	他 16 対

表 3: T4. で用いられた検索クエリ

対立軸	作業 1	作業 2	作業 3
(a)	健康のために飲酒をする	飲酒 健康 適量	飲酒 健康 良い 悪い
(b)	少量 飲酒 実験結果	飲酒 健康 少量	飲酒 少量 健康 良い
(c)	少量 飲酒 心臓病	飲酒 心臓病 適量	飲酒 心臓病 リスク
(d)	少量 飲酒 糖尿病	適量 飲酒 脳 良い 悪い	飲酒 ボケ 防止
(e)	少量の飲酒は良く寝れる	飲酒 糖尿病のリスクを低下	飲酒 寿命 のばす

る言明（対立言明）や、要約対象となる文書集合などはシステムが自動的に推定する必要がある。それゆえ、それらの推定に関する情報も調停要約コーパスに含むように作成作業を設計した。作成作業は表 1 に示す 10 段階で行うこととし、各作業者の作成過程を統制するために表 1 の順番に従って作業を進めることとした。紙面の都合上、各段階の作業を詳細に説明することができないため、4 章で述べる分析と密接に関連する部分のみを以下に説明する<sup>2</sup>。

最初に、各作業者には「ディーゼル車は環境に良い」といった着目言明と、着目言明をクエリとして検索した上位 250 文書が初期文書集合として与えられる<sup>3</sup>。検索エンジンには、自然文を検索クエリとして入力できる TSUBAKI[8] を利用することとした。また、調停要約作成作業の負担を軽減すると同時に、各段階の作業ログを自動的に記録する調停要約作成支援ツールを開発し、各作業者は当該ツール上で作業を行った。

T1. では、対立する言明を公平な視点から理解できるように、各作業者は着目言明のトピック<sup>4</sup>においてどのような論点が存在し、各論点においてどのような意見や根拠が存在しているか調査を行う。その後、T2. では、与えられた文書集合から着目言明と対立する内容の記述を抽出する。ここで抽出対象となる対立軸は 1 つとは限らない。例えば「ダイエット」に関する文書集合においては「痩せる 太る」という軸の他にも、美容観点の「美しい 醜い」、医療観点の「健康 病气」といった軸が考えられる。したがって「ダイエットする」と対立する内容として「太る」「醜い」「病気になる」といった記述を全て抽出するようにした。

T3. では、T2. で抽出された記述が属する対立軸を網羅的に設定し、各対立軸に「ディーゼル車は環境に良い ディーゼル車は環境に悪い」といった形式のラベルを付与する。ただし、T4. 以降の作業者間の比較を容易にするため、3 つの対立軸に関しては、事前に

我々が初期文書集合を調査した結果に基づいて設定し、初期文書集合と共に作業者に与えている。各作業者は、対立軸ごとに両立可能となりそうな状況を記述した後、調停要約を作成できそうな対立軸を、与えられた 3 つの対立軸の他に 2 つ選択し、計 5 つの対立軸に対して調停要約を作成することとした<sup>5</sup>。各作業者によって設定、選択された対立軸がどの程度共通しているか、また、選択された対立軸にはどのような特徴があるかを分析することで、システムが提示すべき対立言明の推定方法を調査する。

初期文書集合は調停要約として適切な記述を必ずしも含んでいるとは限らないため、T4. において、調停要約の記述を含むような文書集合を任意のクエリを用いて検索し、初期文書集合に加えることとした。すなわち、この段階で要約対象となる文書集合が確定し、作業者ごとに差異が現れることとなる。具体的には、T3. で選択した対立軸ごとに、TSUBAKI での検索結果から、調停要約の対象となる文書集合を求めるのに最適と思われるクエリを 1 つ決定し、そのクエリによる上位 50 件の文書を初期文書集合に加える。したがって、5 つの対立軸で 250 件の文書が加えられることになるが、重複する文書の存在があるため、要約対象となる文書数は最終的に 500 弱となる。T4. で用いられたクエリを分析することで、調停要約の対象として適切な文書集合を検索するためのクエリ生成方法を調査する。

T5. では、調停要約として適切な記述を 1 つのパッセージ（調停パッセージ）として抽出する。今回の作業では、我々が直接調停要約と定義する、1 つのパッセージで両立可能となる状況を明確に説明するタイプの調停要約を対象としている。したがって、調停要約の一部として必要な記述ではあるが、その記述だけでは両立可能であることを明確に伝えられない記述は調停パッセージとして抽出しない。直接調停要約は、システムが抽出したパッセージをそのまま利用者に提示するため、そのパッセージを記述をした人物の考えなどがシステムによって捻じ曲げられる危険性が比較的小さい。その一方で、調停要約として最適な表現かつ過不足ない内容となっている保証はない。それゆえ、T8. において、各作業者が理想とする調停要約を自由記述で作成し、抽出された調停パッセージとの差を分

<sup>2</sup>我々はこれまで文献 [2] 等において、情報信憑性判断支援のための分析・評価用コーパスの作成を行っており、調停要約コーパスの作成においても基本的な流れなど重複する部分が多い。重複部分の説明に関しては、それらの文献を参照されたい。

<sup>3</sup>作業者の裁量で初期文書集合を決定することが望ましかったが、作業者の負担軽減や、作業内容の比較の容易さといった理由から初期文書集合を与えることとした。

<sup>4</sup>本稿でのトピックとは、例えば「ディーゼル車は環境に良い」であれば、ディーゼル車に関する話題の内、排出される有害物質、燃費、騒音など環境に影響を及ぼしそうな話題を指す。

<sup>5</sup>対立軸を 5 つに限定したのは労力軽減によるものである。

表 4: T8. で作成された自由記述による調停要約

作業 1	飲酒が健康に良いかどうかは、飲酒量によって事情が変わる。適量の飲酒は、狭心症や心筋梗塞などの心臓疾患を予防する効果のある善玉コレステロールのレベルを上げて、逆に悪玉コレステロールを減らす働きがある。具体的には、赤ワインには血管を保護する作用を持っているポリフェノールが含まれており、白ワインには、抗菌力やミネラルが含まれている、などがある。しかし、過度の飲酒は肝硬変、膵炎、胃潰瘍、痛風などはもちろん、心臓に対しても高血圧、不整脈、心筋症などが起こる可能性がある。さらに、二次的に脳卒中、心不全、突然死などにつながることも少なくない。一方で、飲酒は病気の種類によっても事情が変わる。アルコールには末梢血管を拡張させる働きがあるために、少量の飲酒では血行が良くなり、血圧が下がる。また、善玉コレステロールが増加して、動脈硬化も防ぐと言われている。心筋梗塞や脳梗塞や冠動脈疾患にも良い影響を与えるようである。しかし、少量の飲酒さえも肝臓疾患やガン（喉頭がん、食道がん、胃がん、肝がん、乳がん）には良い影響を与えないようである。また、脳が縮小するという事例もある。したがって、飲酒は飲酒量と病気の種類によって健康に良いか悪いかが変わる。
作業 2	飲酒が健康に良いか悪いかは、飲酒量、飲酒する人に依存する。適量の飲酒であればさまざまな病気のリスクを軽減し、ストレスを軽減し、死亡率を下げるため飲酒は健康に良いと言える。しかし、多量の飲酒はさまざまな病気のリスクを増大させるだけでなく、アルコール依存症の危険性を高めてしまう。また、妊娠、授乳中の飲酒は胎児の健康に悪影響を与えてしまい、未成年の飲酒は脳の発達に悪影響を与えてしまう。さらに薬を服用中の人が飲酒した場合予期せぬ副作用が起きてしまう可能性もある。それゆえ、多量または飲酒する人によっては飲酒は健康に悪いともいえる。
作業 3	飲酒が健康に良いか悪いかということは、飲酒量に依存します。少量から適度の飲酒習慣は、さまざまな病気の予防や、ストレス解消の効果があるなど肉体的にも精神的にも健康だと言われていますが、一方で過度の飲酒習慣は、アルコール依存症の発症や、肝機能障害を引き起こすことがあるため、健康に悪いと言えます。

析するために用いる。T9. では、抽出されたパッセージがどの対立軸の調停要約となっているかを分類し、T10. では、対立軸ごとに分類されたパッセージに対して調停要約としての適切性の観点から線形順序でのランキングを行う。また、各パッセージに対し T8. で作成した理想の調停要約との内容や表現などの近さを総合的に判断して最高の A 評価から最低の D 評価までの絶対評価を行う。

情報信憑性判断という観点からは、調停要約として提示された記述が、どのような人物や組織から発信されたものであるかという情報も重要であると考えられる。それゆえ、我々は情報発信者抽出に関する研究 [3] を行っている。T6. と T7. は情報発信者に関する作業を行っているが、紙面の都合により説明を割愛する。

## 4 調停要約コーパスの分析

今回の作業では、1 つの着目言明に対して 4 名の作業者を割り当て、4 通りの調停要約に関する情報を付与している。調停要約コーパスには、最終的に、6 つの着目言明に対する計 24 通りの調停要約に関する情報が収録される予定であるが、本稿執筆時点では調停要約作成作業が継続中である。それゆえ、本稿では、3 名の作業者が作業を完了している「飲酒は健康に良い」を着目言明とした場合の調停要約に関する情報に基づいて定性的な分析を行う。なお、作業者は情報工学を専攻する大学生および大学院生である。また、T1. から T10. までの作業にかかった時間は、作業 1 が 39 時間 15 分、作業 2 が 36 時間、作業 3 が 30 時間 10 分であった。

本稿では以下の 3 点について分析する。1 点目は T3. において各作業者が選択した対立軸に関して、2 点目は調停要約の対象となる文書集合を決定するために T4. で用いられた検索クエリに関して、3 点目は T5. で抽出された調停パッセージと T8. で作成された自由記述による調停要約との差に関してである。

各作業者が T3. で選択した対立軸を表 2 に示す。(a) から (c) は初期文書集合と共に与えられた作業者共通の対立軸であり、(d) と (e) が各作業者が任意に作成した対立軸から調停要約を作成できそうな対立軸として選択された軸である。各作業者が選択した (d) と (e)

の対立軸は、他の 2 名の作業者においても対立軸として設定されており、対立軸の認識に関しては作業者間のズレは小さいと考えられる。しかしながら、作業 1 の (d) と作業 2 の (e) を除いて複数の作業者が共に選択している対立軸は存在しないため、対立軸の重要性に関してはズレがあると思われる。T2. で抽出された記述が 1 つの対立軸あたりに属する数は、設定された対立軸全体での平均が 29.8 であったのに対し、選択された対立軸での平均は 92.2 であった。したがって、属する記述の数によって対立軸の重要性を計算することができると考えられる。

次に、各作業者が要約対象とした文書集合の重複度合いを図 1 に示す。全作業者に共通の 251 文書の内、250 文書は初期文書集合であるため、T4. において追加された文書集合において全作業者に共通する文書数は 1 であり、検索された文書集合はほとんど重複しなかった。また、T10. において B 評価<sup>6</sup>とされた調停パッセージの数は 3 名の合計で 42 パッセージであったが、その内の 22 パッセージは各作業者が検索した文書集合から抽出されたものであった。したがって、要約対象とする文書集合の決定は調停要約の精度に影響する重要な処理であり、文書集合を決定するための検索クエリも重要な要素である。各作業者が T4. で用いた検索クエリを対立軸ごとに整理したものを表 3 に示す。表 3 の対立軸の記号は表 2 の記号に対応している。TSUBAKI が自然文で検索可能であることは各作業者也理解しており、T4. において 3 名の作業者が調査した計 57 クエリの内、22 クエリは自然文でのクエリであった。しかしながら、初期文書集合に加えた文書集合の検索に用いたクエリは表 3 に示すように単語列であるものが多かった。この結果について各作業者に質問したところ、「最初に自然文で入力したが、思うような文書が検索されなかったため単語列で検索した」という回答であった。この原因として「飲酒 健康 良い 悪い」のように、良い面と悪い面の両方を記述している文書を検索するという調停要約特有の要求を満たすクエリを文の形式で表現しにくかったことが考えられる。以上から、調停要約として適切な記述

<sup>6</sup>理想の調停要約との近さを尺度とした絶対評価のため、本稿においては A 評価とされた調停パッセージは存在せず、B 評価が最高評価となっている。

表 5: T5. で抜粋された調停パッセージ

作業 1	酒（リンク 2 参照）は百薬の長と言われます。日本人男性を対象とした調査の結果、2 日でビール大瓶 1 本、もしくは日本酒 1 合程度飲酒する者の死亡率が最も低いことが報告されています。適量飲酒ならば、健康に好ましい影響を与えそうですが、それでは適量飲酒とはどれくらいでしょうか。アルコールが体内に入ると、アセトアルデヒドという物質に変化します。これが、顔を赤くする原因物質です。アセトアルデヒドはアルデヒド脱水素酵素 2（ALDH2）により分解されますが、日本人の 1 割程度の方では ALDH2 活性がまったくなく、少量の飲酒でアセトアルデヒドが急上昇します。このような方では、奈良漬を食べただけで顔が赤くなります。また、4 割程度の方は、酵素活性が弱いのでビールをコップ一杯程度飲んだだけで顔が赤くなります。アセトアルデヒドは、悪心、動悸などの不快症状の原因となります。また、飲酒に関連するガン（食道ガン、大腸ガンなど）の発症に寄与していると考えられています。そのため、ALDH2 活性の弱い人は言うに及ばず、ALDH2 活性に問題がない方でも、酒の量が過ぎればアセトアルデヒドが発生し、長期的にはガンなどの危険が高まります。従って、酒は万人にとって必ずしも薬とは言えません。
作業 2	適度の飲酒をしている男性は禁酒もしくはほとんど飲まない人と比べて 1.27 倍、女性は禁酒をしている人と比べて 2 倍以上も、健康的と報告されました。この結果、適度の飲酒をしている人は、禁酒をしている人・あまりお酒を飲まない人・酒量が多い人と比べて標準以上の健康を保っているようです。また、この研究者によると、適度の飲酒は必ずしも危険ではなく、むしろ健康にはいい場合もあり、動脈硬化といった心疾患や、脳梗塞を防ぐ手助けをしてくれる可能性もあるとのこと。ただし、深酒や過度の飲酒となると健康にはよくないので、やはり飲酒をする場合は“ほどほど”がお勧めのようです。
作業 3	古くから云われている「酒は百薬の長」という言葉。適量のお酒が健康によいというのには、科学的裏付けがあるようです。これは「適量の飲酒をしている人の死亡率が、まったく飲まない人よりも低い」という、米国保健科学協議会のレポートに基づいています。もちろんいうまでもありませんが、過度の飲酒をする人の死亡率が、飲まない人より圧倒的に高いことは、認識しておいていただく必要があります。さて、適量の飲酒が健康によいという理由は、ひとつにはストレス発散の効果が精神にいい影響を及ぼすということもありますが、他に心臓疾患予防に具体的な効能があるらしいのです。アルコールが、狭心症や心筋梗塞などの心臓疾患を予防する効果のある善玉コレステロール（HDL コレステロール）のレベルを上げて、逆に悪玉コレステロール（LDL コレステロール）を減らす働きがあるとのこと。また、先に述べたストレス発散の効果は、ストレスに反応する冠動脈発作を減らす働きがあるといわれています。ただし、繰り返しになりますが、これはあくまで「適量」の範囲内でのこと。過度の飲酒が健康によくはないことは、皆さん、肝に銘じておきましょう。

を含む文書を検索するという観点からは、単語列を用いた方が適している可能性が示唆された。ただし「飲酒 糖尿病のリスクを低下」のように単語と句を組み合わせたクエリも存在したことから、必ずしも単語列が最適というわけではない。この点に関する分析は今後の課題である。

表 4 は「飲酒は健康に良い 飲酒は健康に悪い」の対立軸に対して T8. で作成された理想の調停要約であり、表 5 は同じ対立軸において T10. のランキングで 1 位となった調停パッセージである。理想とする調停要約に関しては、作業 1 が「病気の種類」、作業 2 が「飲酒する人の状態」という観点からも記述しているが、基本的には 3 名とも「飲酒量」という観点からまとめており、多くの人に共通する調停要約の観点が存在するように思われる。また、表 4 の調停要約と表 5 の調停パッセージを比較した場合にも、細かな違いはあるものの「飲酒量」などの観点は共通しており大意において差はないと思われる。したがって、1 つのパッセージを抽出して提示する直接調停要約の考え方に大きな問題はないと考えられる。しかしながら、理想とする調停要約と完全に同一というわけではないため、その差を改善するために今後さらに分析を行う必要がある。

## 5 おわりに

本稿では、調停要約コーパスの作成作業を説明し、コーパスを分析した結果について報告した。分析は、各作業者が選択した対立軸、調停要約の対象となる文書集合を決定するために用いられた検索クエリ、抽出された調停パッセージと自由記述による調停要約との差の 3 点に関して行った。今後は、他の着目言明に対する調停要約に関しても分析を行い、それらの結果と合わせて調停要約自動生成システムに反映したいと考えている。

## 謝辞

本研究は、独立行政法人情報通信研究機構の委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」プロジェクトの成果である。

## 参考文献

- [1] H. Shibuki, T. Nagai, M. Nakano, R. Miyazaki, M. Ishioroshi, and T. Mori, A Method for Automatically Generating a Mediatory Summary to Verify Credibility of Information on the Web, *Proc. of the 23rd International Conference on Computational Linguistics*, pp.1140–1148, 2010.
- [2] M. Nakano, H. Shibuki, R. Miyazaki, M. Ishioroshi, K. Kaneko, and T. Mori, Construction of Text Summarization Corpus for the Credibility of Information on the Web, *Proc. of the 7th Language Resources and Evaluation Conference*, pp.3125–3131, 2010.
- [3] R. Miyazaki, R. Momose, H. Shibuki, and T. Mori, Using Web Page Layout for Extraction of Sender Names, *Proc. of the 3rd International Universal Communication Symposium*, pp.181–186, 2009.
- [4] 衛藤純司, 奥村学. 文書横断文間関係タグ付コーパスの構築. 言語処理学会第 11 回年次大会, pp.482-485, 2005.
- [5] D. R. Radev, A common theory of information fusion from multiple text sources, step one: Cross-document structure. *Proc. of 1st ACL SIGDIAL Workshop on Discourse and Dialogue*, pp.74–83, 2000.
- [6] 村上浩司, 増田祥子, 松吉俊, 乾健太郎, 松本裕治, 言明間の意味的関係の体系化とコーパス構築, 言語処理学会第 15 回年次大会, pp.602–605, 2009.
- [7] 村上浩司, 水野淳太, 後藤隼人, 大木環美, 松吉俊, 乾健太郎, 松本裕治, 文間意味的関係認識による言論マップ生成, 言語処理学会第 16 回年次大会, pp.559–562, 2010.
- [8] K. Shinzato, T. Shibata, D. Kawahara, C. Hashimoto, and S. Kurohashi, TSUBAKI: An Open Search Engine Infrastructure for Developing New Information Access Methodology, *Proc. of the Third International Joint Conference on Natural Language Processing*, pp.189–196, 2008.