

## 要約筆記品質評価システムにおける筆記者評価

### Discriminate Writer Type of Quality Evaluation System of Summary Transcript

高尾 哲 康

Takao Tetsuyasu

#### 1. はじめに

聴覚障害者や高齢者への情報保障手段のひとつに要約筆記がある。要約筆記には「PC 要約筆記」と「手書き要約筆記」があり、いずれも要約筆記者が講演や番組などを聞き取り、リアルタイムで要約を行ない、キーボードや手書きで入力する。一般に日本語の発話速度は200～400文字/分であり、要約筆記者による入力量はPCの場合で100～200文字/分、手書きの場合で40～80文字/分となっている。要約筆記者は「速く」、「正確に」、「読みやすく」の3原則をもとに、技術の向上を目指してさまざまな研修プログラムで訓練を重ねる。個々の研修プログラムでは要約筆記の品質の尺度として、要約筆記利用者からのフィードバックや意見・要望を受けることが多い[1]。これらのフィードバックは個々の事例として受けることが多く、定量的な品質評価を受けることはほとんどなかった。そのため、長期間の研修を経ても要約筆記の品質向上の実感が得られにくくなっていた。これまで筆者らは講演者の発話内容のテキストと要約筆記者が入力したテキストをもとに定量的な評価ができるシステムを試作した[2][3][4][5]。評価計算には重み付き編集単位要素（主に形態素原形）の編集距離（Edit Distance または Levenshtein Distance）に基づく要約評価方式を提案した。しかし、この方式では局所的範囲での評価には有効性が確認できたものの、全体としての評価である要約筆記者評価は個々の評価の単なる積み上げになっているだけで評価がなかなか見えにくいという欠点があった。また、要約筆記されたテキストどうしの関連性がわからない、要約筆記者それぞれのくせがわかりにくいなどの問題もあった。

本論文では、講演者の発話内容のテキストと要約筆記者が入力したテキストをもとに定量的な評価を行なうシステムにおいて、より全体的な評価、すなわちそれぞれの要約筆記者のタイプやくせなどの評価ができることを検討した。評価手法には、情報検索や文書分類などの分野で利用されつつある統計的手法として、多次元尺度法（MDS: Multi-Dimensional Scaling）と潜在意味解析（LSA: Latent Semantics Analysis、潜在的意味インデキシングともいう）を利用した[5]。これにより、局所的評価だけでなく全体的な評価がより正確に行なえるようになり、要約筆記の技術レベル診断や改善点など、今後の研修プログラムの指針となる可能性が開ける。また、コンピュータによる自動要約システム[6]への知見が得られると考える。

デジタル放送時代に入り、字幕放送が広く認知されてきている。しかし、現在の字幕放送はリアルタイム字幕放送であっても話者の発言をほとんどすべて文字化している。このため、字幕テキストをととても読み切れない、そのために映像を楽しむことができないという意見が出てきている。脳科学研究の成果によれば、言語においては視覚より聴覚のほうが時間当たりの情報量を多く処理できると言われている。そのため、番組にもよるが、発言やナレーションを要約して字幕表示することにより、多くの人にとってより読みやすくなる場面が出てくると考える。本研究では正確で読みやすいコンパクトな要約をするための指針となることを目指している。

## 2. 品質評価に利用した要約筆記データ

要約筆記講習会の研修プログラムで使用した発話テキスト（Sと表わす）と要約筆者6名が手書きでリアルタイム要約筆記したテキスト（K1～K6と表わす）を利用した。詳細を表1に示す。発話テキストには要約筆者の存在を意識してゆっくりと朗読したエッセイ文（約6分）を利用した。表1には、これまでに開発したシステム[2][3][4]による、編集距離に基づく要約評価結果も参考として載せた。文字数には句読点や記号（矢印記号「→」、項目を表わす中黒「・」）、繰り返し記号（「〃」など）、削除記号（訂正線）などを含めている。手書き要約筆記の場合はPCによる要約筆記とは異なり、二次元的な表現や複数行にわたる括弧記号や複数行にまたがる矢印記号、横方向の波括弧付きの挿入文字、取り消し線付き文章などが含まれていることがある。そのような場合には、計算機可読テキストにする際にはその意図がなるべくわかるようにXMLタグ付きテキストにした。

表1. 要約筆記テキストと要約評価

	文字数	文字数 /分	要約率 (%)	全発話との比較		要約 評価
				文字列	形態素	
全発話(S)	1808	226.0	-			
筆者1(K1)	545	68.1	30.14%	0.5210	0.6100	0.5687
筆者2(K2)	434	54.3	24.00%	0.4426	0.5443	0.4938
筆者3(K3)	539	67.4	29.81%	0.4316	0.4216	0.4342
筆者4(K4)	665	83.1	36.78%	0.6354	0.6739	0.6591
筆者5(K5)	395	49.4	21.85%	0.2809	0.2810	0.2663
筆者6(K6)	443	55.4	24.50%	0.3385	0.3403	0.2870

## 3. 要約筆記品質評価システム

要約にはさまざまなレベルがある。「テープ起こし」（要約率 90%）や丁寧な要約筆記である「概要要約筆記」（要約率 40～60%）などがある。PCを利用した2人連携要約筆記では、後編集

処理をほどこすだけでも十分概要要約筆記のレベルに到達できる。本システムでは、PC や手書きを問わず、さまざまな要約レベルにも対応できるようにしている。本システムの概要、システム構成を図1に示す。

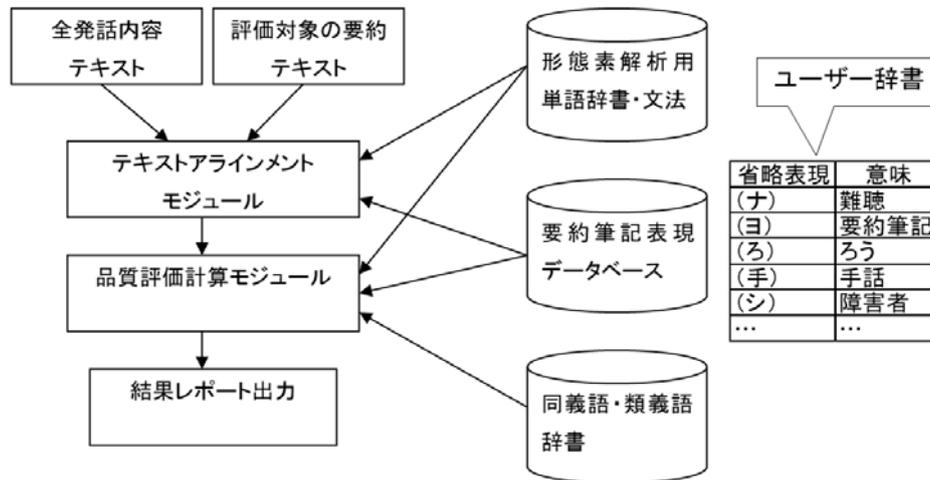


図1. システム構成図

本システムはテキストアラインメントモジュールと品質評価計算モジュールから構成される。テキストアラインメントモジュールは全発話テキストとその要約テキストを入力し、統計情報と言語情報をもとに、動的計画法を利用して対応する文や段落を関連づけるモジュールである（m文対n文の対応付け）[7]。品質評価計算モジュールは、表記のゆれ（漢字の読みのひらがな・カタカナ表記、異字体、異表記など）や要約筆記特有の省略表現などを吸収して正規化した形態素解析結果のテキストに対し、単語コスト、品詞コスト、単語間接続コスト、重複出現コスト（本システムでは出現のたびに単調減少することで計算）をもとに、重み付き編集単位要素（主に形態素原形）の編集距離に基づき、要約の品質評価計算を行なう[3]。要約筆記者評価についてはこの品質評価モジュールにおいて、文字列と形態素列の N-gram に基づいた多次元尺度法と潜在意味解析の手続きを追加したことで行なえるようになっている。評価対象はテキスト全体およびアラインメント単位で行なう。また、各種の統計処理には、統計解析のためのフリーソフト R およびそのライブラリ群を利用した。

多次元尺度法は、個体間の類似の度合いや非類似度の度合いを、それぞれの個体を測定して得られる多変量の数値をもとに個体間距離を計算することで求め、類似性の高いものが近くに配置するようにする方式である。その結果を2次元あるいは3次元に描画することで個体間の関係を探ることができる。

潜在意味解析は単語の持つ多義性と多様性に対処した統計的技法であり、多変量のゆえに高次元となる単語文書行列を低次元に圧縮して近似的に表現する。これにより関連性の強い単語どう

しがより強調されたものとなり、文書間の意味的類似度が明確になる。つまり、文書どうしの分類を効果的に行なうことができる。

#### 4. 実験結果

要約筆記者評価計算には、解析対象テキストを文字列とみた場合と形態素列とみた場合のそれぞれについて N-gram を求め、これらに対して多元尺度法と潜在意味解析処理を行なった。形態素列の場合は N=1 のユニグラム (Unigram、単語頻度) とバイグラム (Bigram、2 単語接続頻度) を利用した。さらに、形態素列の場合は助詞や助動詞などの機能語を排除して内容語のみに限定する集計を行なった。

多元尺度法では、個体間の距離座標値を求める方法として計量的方法と非計量的方法がある。本実験では、単語頻度や単語接続頻度の数値から個体間の距離を計算するので、非計量的方法に基づいて各文書の座標を求めた。バイグラムから各テキストの座標を求め、図示した結果を図 2 に示す。K4 が最も全発話テキストに近いことがわかる以外はあまり傾向が見いだせていない。クラスター分析を行なえば、個体間距離の計算方法やクラスター作成方法によってはうまく分類できる場合もある。

潜在意味解析では、ユニグラムをベースに TF (Term Frequency、索引語頻度)・IDF (Inverse Document Frequency、文書頻度逆数) の重み付けと正規化を行なった単語文書行列をもとに特異値分解によって次元を圧縮している。なお、特異値分解は潜在意味解析以外では対応分析などで用いられている。3次元に圧縮して各テキストの座標値を求めた結果を図 3 に示す。S や K1~K6 は多元尺度法の場合と同じである。これにより、やはり K4 が S に最も近いので、全発話内容に最も近いことがわかる。また、K1 と K2、K5 と K6 が近いことから、テキストに共通するくせが現われていることもわかる。一例を挙げると下記のようなになる。

「しゃべることよりも聞くことのほうが倍も重要だということで人間に耳を 2 つつけられたということです」(S)

「しゃべるより聞くことが重要と耳を 2 つ与えられた」(K1)

「しゃべるより聞くことがより重要だと耳を 2 つ与えられた」(K2)

「聞くことのほうが話すことよりも重要だ」(K5)

「よりよく聞き、話すことのほうを半分に」(K6)

この例では、「しゃべる」→「話すこと」と、話し言葉向けの表現から書き言葉向けの表現に改められているというくせが表われている。他にも、助詞の使い方、読点の使い方（こちらは発話者の原稿と比べて）などに要約筆記者のくせが表われることがわかった。

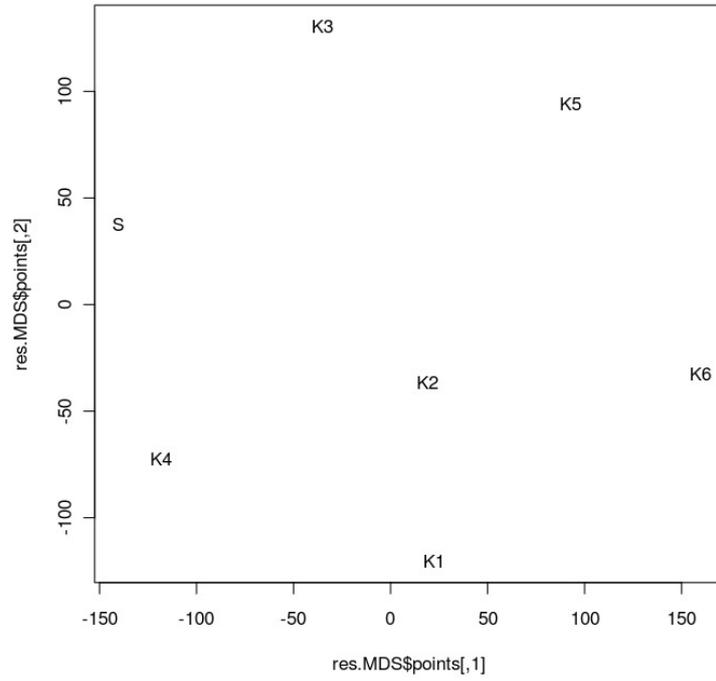


図2. 多次元尺度法による解析結果

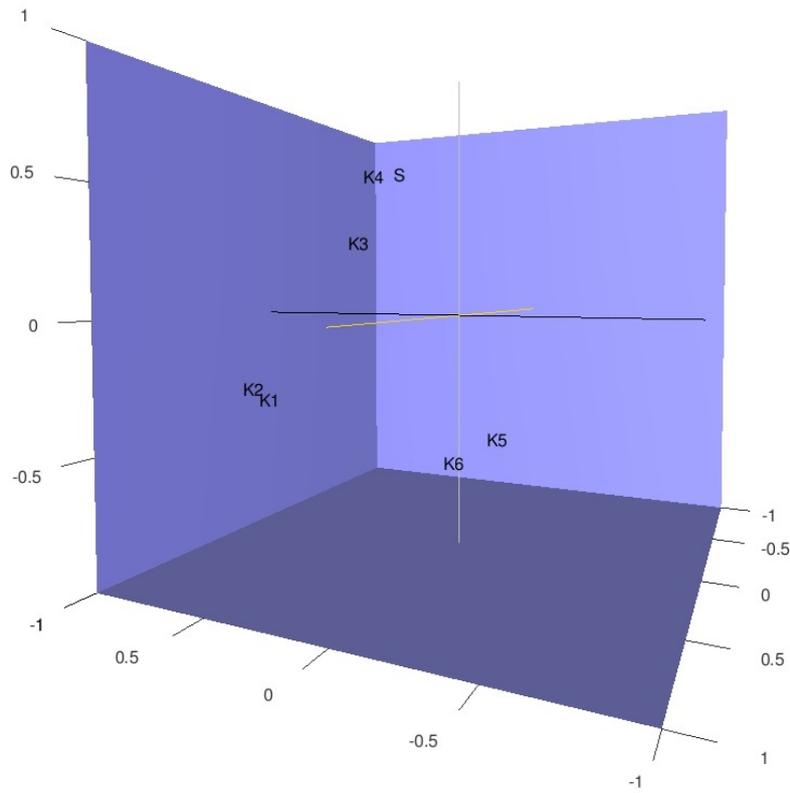


図3. 潜在意味解析による解析結果

## 5. まとめ

本実験から要約筆記者評価を行なうシステムの有効性がある程度確認できた。特に潜在意味解析により、要約筆記者のタイプやくせを見つけ出すことが容易になった。今後は、さまざまな要約筆記者例に適用し、従来の局所的評価と今回の要約筆記者評価とを統合した詳細な分析を進めるとともに、品質評価計算手法の改良やパラメータの最適化などを行なっていく。

## 参考文献

- [1]要約筆記再履修資料、名古屋市登録要約筆記者の会編 (2007)
- [2]高尾哲康、要約筆記の品質評価の試み、IPSJ72 全国大会、5G-3、(2010)
- [3]高尾哲康、要約筆記の品質評価システム、情報科学技術フォーラム FIT-2010、7D-7、(2010)
- [4]高尾哲康、要約筆記品質評価システム富山国際大学現代社会学部紀要、Vol.3、2010
- [5]高尾哲康、要約筆記品質評価システムの改良、FIT2011、3Q-5、(2011)
- [6]特集 テキスト自動要約、情報処理、Vol.43、No.12 (2002)
- [7]高尾哲康、対訳テキストコーパスからの対訳語情報の自動抽出、情報処理学会自然言語処理研究会、115-8、(1996)