

音声認識との自動連携による要約筆記の品質向上

高尾 哲康[†]富山国際大学現代社会学部経営情報専攻[‡]

1. はじめに

聴覚障害者や高齢者への情報保障手段である PC 要約筆記では、要約筆記者が講演などを聞き取り、必要に応じて要約を行ない、キーボードから入力する。要約筆記者は「速く」、「正確に」、「読みやすく」の 3 原則をもとに、さまざまな研修プログラムで訓練を重ねる [1]。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評価を行ない、要約筆記者支援としてよりよい要約筆記表現を抽出する機能をもつシステムを試作した [2]。現在の PC 要約筆記が IPtalk [4] などを利用してネットワーク経由でひとつの発話文の前半と後半などに分けて分担入力し、2~4 人連携で行なわれている実情に合わせ、複数人による要約筆記文を自動連携することにより、要約筆記者の労力軽減とともに要約筆記文の品質向上をめざす試みを行なった [3]。今回、実利用が十分可能となってきた音声認識システムを活用し、要約筆記と組み合わせて高品質な要約テキストを出力する実験を行なった。

2. 要約筆記データと音声認識システム

要約筆記研修プログラムで使用した発話テキスト(観光ガイド約 4 分、T1 で表わす)と PC 要約筆記者 4 名による要約筆記テキスト(P1~P4 で表わす)を利用した。詳細を表 1 に示す。要約筆記者 2 人連携の部分は P1~P4 の各要約筆記テキストをもとに、2 名ずつ自動連携した結果である。

音声認識システムは入手が容易な AmiVoice SP2、ドラゴンスピーチ 11J、Julius-4.3.1、Siri (MacOSX)、UD トーク (Powered by AmiVoice)、VoiceRepPro (Google)、Windows 音声認識の 7 システムを利用した。マイク入力以外にライン入力または音声ファイルからの認識が可能である。講演者ワイヤレスマイクからの電波を広帯域レシーバで受信したり、IC レコーダや屋外での観光ガイド用イヤホンクリボ (CleVo)、指向性マイクなどから品質のよい音声を取り込んで認識できる。多くのシステムで IPtalk [4] のテキスト

入力域やエディタ画面などに認識結果を直接入力可能である。要約筆記の場合と同じ音声データを認識した結果の要約評価を表 1 に R1~R7 で示す。要約筆記との連携にはトレーニング(音響学習、主に固有名詞登録)前の認識率が高く、トレーニング作業も容易なシステム(表 1 の R1、トレーニング後を R1T と表わす)を対象とした。

3. 要約筆記品質評価システム

本システムはテキストアライメントモジュールと品質評価計算モジュール、複数テキスト連携モジュールから構成される。テキストアライメントモジュールは発話テキストと要約筆記や音声認識テキストを入力とし、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)。品質評価計算モジュールは、表記ゆれ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(MeCab を利用)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(出現のたびに単調減少)を統計処理することにより、要約の品質評価(要

表 1. 音声認識と要約筆記と連携による要約評価

		文字数	入力速度 (文字数/分)	要約率 (%)	要約評価
発話者	T1	972	249.2		
音声認識	R1	904	231.8	93.0%	0.8069
	R1T	944	242.1	97.1%	0.9258
	R2	901	231.0	92.7%	0.8092
	R3	981	251.5	100.9%	0.5165
	R4	831	213.1	85.5%	0.7634
	R5	940	241.0	96.7%	0.7984
	R6	811	207.9	83.4%	0.6777
	R7	857	219.7	88.2%	0.6640
要約筆記者	P1	533	136.7	54.8%	0.6522
	P2	492	126.2	50.6%	0.6052
	P3	370	94.9	38.1%	0.4902
	P4	497	127.4	51.1%	0.6180
要約筆記者1人 +音声認識	P1+R1	882	226.2	90.7%	0.8724
	P2+R1	812	208.2	83.5%	0.8558
	P3+R1	877	224.9	90.2%	0.8445
	P4+R1	931	238.7	95.8%	0.8287
	P1+R1T	925	237.2	95.2%	0.9261
	P2+R1T	871	223.3	89.6%	0.9051
	P3+R1T	890	228.2	91.6%	0.9001
	P4+R1T	951	243.8	97.8%	0.8904
要約筆記者2人連携	P1+P2	678	173.8	69.8%	0.7320
	P1+P3	669	171.5	68.8%	0.7130
	P1+P4	744	190.8	76.5%	0.7259
	P2+P3	629	161.3	64.7%	0.6744
	P2+P4	708	181.5	72.8%	0.7113
	P3+P4	649	166.4	66.8%	0.6752

Speech Recognition Assist Function of Quality Improvement Support System of Summary Transcript

[†]Toyama University of International Studies

Faculty of Modern Society

[‡]Tetsuyasu Takao

約評価)の計算を行なう[2][3]。複数テキスト連携モジュールは、複数の要約筆記文をマージする機能をもち、品質評価計算モジュールのアルゴリズムを流用することで実現している。

複数文のマージは2つの文の相互に異なる部分を抽出し、合成することで行なう。表2に音声認識結果の文「西郷さんは軍曹対象として江戸に来た」と要約筆記の文「官軍総大将として江戸へきた」をマージする例を示す。各文の形態素列において、コストは形態素解析用単語辞書に格納されている形態素コストを初期値としている。各形態素を0~1の重み付き編集単位要素とみなして編集距離を求める。編集距離とは列Aと列Bについて、Aを編集操作(削除、挿入、置換)してBにするときの必要最低限の操作数のことである。評価値は編集操作コストを2つの文の形態素コスト値の総数で割り、数値の範囲を0~1に正規化した数値にした。0に近ければ2つの文の相違が多く、1に近ければ相違が少なくなる。各セル値E_{ij}の計算は表2の式にて全セルについて計算を行ない、表の最右下のセル値を1から引いた値が2つの文の評価値となり、この値が1に近いほど類似度が高いことになる。

2つの文のマージの際の書き換え候補の抽出は次のように行なう。表2の評価値を算出するマトリクスにおいて、最右下のセルから最左上のセルまで評価値が最も小さくなる方向(上方、左方、左上方のいずれか)に順次たどることで2つの文の各形態素の対応セルが求まる。次に、2つの文の対応関係のうち相互にマッチしないもの(前後のセル間で評価値の差が大きい場合)を抽出する。表2の例では、列方向と行方向の文を対応させて、

- ・「西郷さんは」⇔文頭
- ・「軍曹対象として」⇔「官軍総大将として」(相違部分とみなし、要約筆記を優先)
- ・「江戸に」⇔「江戸へ」(同評価値)
- ・「来た」⇔「きた」(同上)

が該当する。マージの結果、「西郷さんは」、「官軍総大将として」、「江戸に」、「きた」の順となり、マージ結果の文として、「西郷さんは官軍総大将として江戸へきた」が得られる。

4. 実験結果

音声認識システムと要約筆記について、単独の場合と連携した場合について、発話文と比較した品質評価結果の要約評価を表1右側および図1に示す。音声認識システムと要約筆記者1名を組み合わせるほうが要約筆記者2名連携の場合よりも要約評価が高くなっている。特にトレーニングを行なった音声認識システムと連携

する場合は要約評価も0.9を越えることもあり、要約筆記者のミスによる悪影響が出やすくなる。

5. まとめと今後の課題

品質のよい音声を得られる環境であれば音声認識システムと要約筆記者の連携による品質向上の効果が高いことがわかった。要約筆記の本来の目的は情報保障にあるので、発話に対するのと同様、少々冗長性や文体の不自然さ、誤りなどに対して寛容性があるのが現実である。今後は、特に固有名詞や数字などは正確さが求められるので、音声認識システムとの最適な役割分担方法などを検討していく。

参考文献

- [1]話しことばの要約、三宅初穂、全国要約筆記問題研究会(2012)
- [2]高尾哲康、要約筆記品質評価システムにおける要約表現抽出、IPSJ76、2F-2、(2014)
- [3]高尾哲康、複数要約筆記の自動連携による要約筆記の品質向上、IPSJ77、5F-5、(2015)
- [4]IPtalk http://www.geocities.jp/shigeaki_kurita/

表2. 評価値計算と複数テキスト連携

	コスト	西郷	さん	は	軍曹	対象	として	江戸	に	来	た
コスト	0.0000	0.2000	0.192	0	0.2649	0.2109	0.3749	0.3020	0	0.145	0
官軍	0.3649	0.1528	0.2784	0.2865	0.2865	0.4393	0.5275	0.5476	0.6732	0.6732	0.6793
総大将	0.3749	0.3098	0.4354	0.4434	0.4434	0.5962	0.6845	0.7046	0.8302	0.8302	0.8363
として	0.479	0.3298	0.4555	0.4635	0.4635	0.6163	0.7046	0.6845	0.8101	0.8101	0.8162
江戸	0.3000	0.4555	0.5811	0.5891	0.5891	0.7419	0.8302	0.8101	0.6845	0.6845	0.6906
へ	0	0.4555	0.5811	0.5891	0.5891	0.7419	0.8302	0.8101	0.6845	0.6845	0.6906
きた	0.431	0.4735	0.5991	0.6072	0.6072	0.7600	0.8482	0.8282	0.7026	0.7026	0.7086
た	0	0.4735	0.5991	0.6072	0.6072	0.7600	0.8482	0.8282	0.7026	0.7026	0.7086

$$E_{i,j} = \min(E_{i-1,j} + C_{i-1}/C, E_{i,j-1} + C_{j-1}/C, E_{i-1,j-1} + A)$$

$$A = \begin{cases} 0 & : i-1 \text{ と } j-1 \text{ の位置の形態素がマッチ} \\ & \text{(表記基本形、品詞、同義語)した場合} \\ (C_{i-1} + C_{j-1})/C & : \text{上記以外}(C: \text{コスト値の総和}) \end{cases}$$

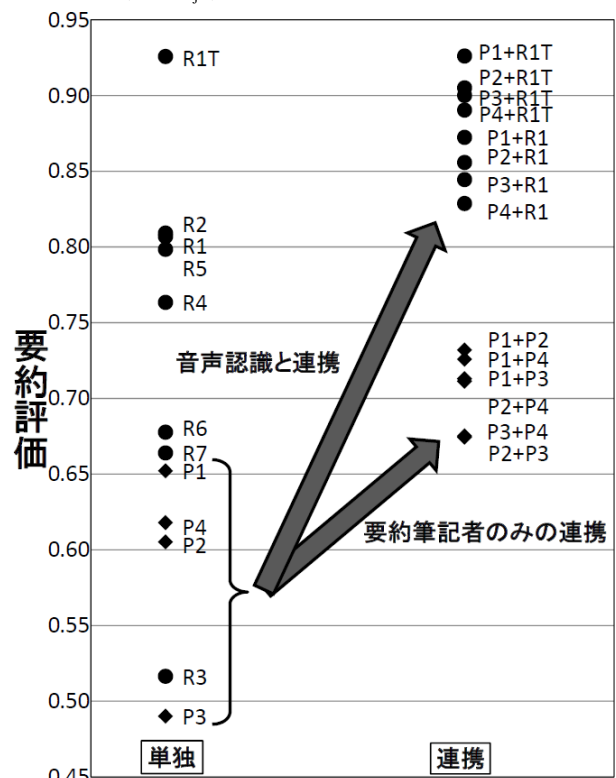


図1. 単独および連携時の要約評価の変化