

視覚障がい者向け音声情報スキミングのための 高速再生技術の高度化および受聴支援装置に 関する研究開発(障害者支援研究開発)

○田澤 直幸、岩鼻 幸男、都木 徹、山崎 順一、富田 豊(NHK-ES)
今井 篤(NHK技研)
鳥原 信一(慶應義塾大学 S.F.C研究所)

発表内容について

- 研究の目的
- 視覚障がい者の情報取得について
- 高速再生アルゴリズムについて
- 高速再生アルゴリズムの外国語への対応について
- 目的別に高速再生アルゴリズムについての検討
- 視覚障がい者による評価実験について
- テキストデータを活用した合成音声の話速変換
- 本機能を実装したアプリ開発など
- まとめ

研究の目的

- 音声の高速再生を活用した、短時間でより多くの情報を取得する技術の開発
- 同じ速度倍率であれば、従来技術に比べて「聴きやすい」と感じさせるアルゴリズムの開発
- 視覚障害者の利用するDAISY図書再生装置やソフトウェアへの実装

視覚障がい者の情報取得について

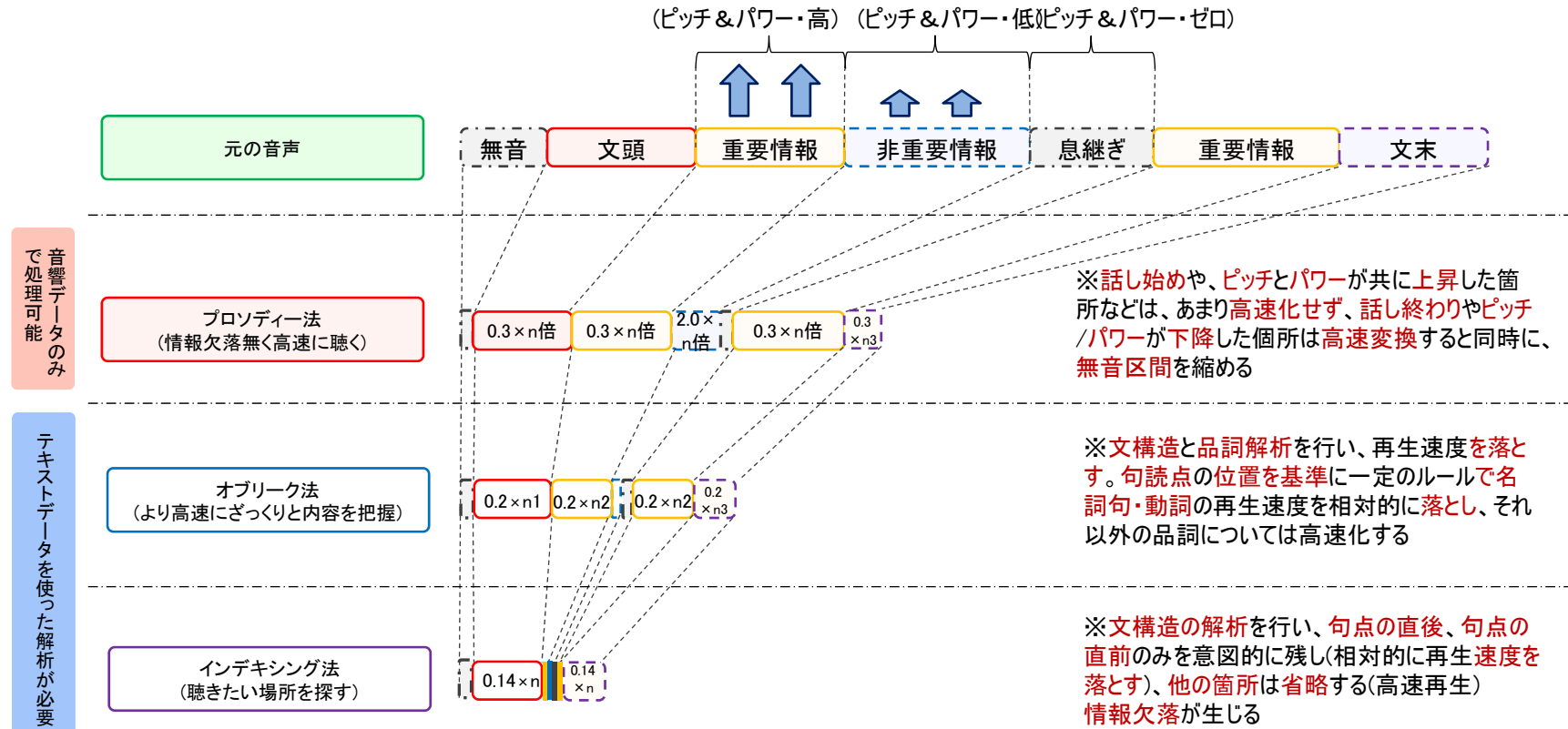
- 視覚障がい者の主な情報取得手段は「音声」
 - ※テレビ・ラジオの音声、
録音図書やスクリーンリーダーなどの活字媒体を音声化したもの
- 音声情報は、最後まで聞かないと文章の全体的な概要が把握できないことが多い
- 多くの情報を短時間で得るために再生速度を上げるが、聴き取り能力の限界付近では聴き取りそのものがつらく、また疲労も増大する
- より高速で聴きたいが、市販機器では最大3倍速程度が限界、音質も様々
- 必要な所だけを聴きくために情報を選別したい
- 音質が損なわれず、高速で聴きやすい音声再生装置(ソフトウェア)が欲しい

高速話速変換アルゴリズム

- NHK放送技術研究所が開発した「話速変換技術」を応用した技術
- 文章の全体を一様の速度倍率で高速化はしない
- 音声データから、重要／非重要箇所を仮定し、適応的に話速を変換する
- 音響処理のみで動作が軽量、リアルタイムで話速の変換が可能
- 辞書などの知識処理が必要なく、言語に依存しない処理が可能

高速話速変換の概念

* 聴き手のニーズを考慮し、効率的情報取得に活用出来る3種類的高速再生方法を検討した。



※1:ピッチ(ピッチ周波数) 音声の有声音区間の基本周波数.

※2:パワー 音声波形に20msec幅のハミング窓をかけたあとの各標本値の二乗和.

高速話速変換の外国語対応(基礎検討)

- 日本語用に開発したアルゴリズムをそのまま外国語に適用した際に得られた意見
 - 単語の一部が欠落してしまう場合があった
 - 発話位置による速度変換は、文章のリズムを変えてしまい聞きづらい
- 外国語に適応する高速話速変換アルゴリズムの検討
- 英・独・仏・西・中の5言語についてネイティブを対象に検討
- 言語の韻律の特徴にあわせ、ピッチとパワーのバランスを調整して伸長倍率を決定する式を作成

日本語と外国語の韻律等の違いについて

	日本語	外国語 (英・独・仏・西・中)
1.アクセントの違い	高低アクセント	強勢アクセント
2.発話リズム	抑揚があり、リズムが一定でない	単語ごとにリズムに法則性がある
3.話速変換パラメータの重み付け	ピッチ重視	パワー重視
4.発話位置による変換	話し始めを遅く	発話位置による変換は行わない

言語ごとの韻律の違いなどを考慮したパラメータ算出式

$RT(t) = W \times (Pi(t)/PiAV) + (1 - W) \times (Po(t)/PoAV)$			
$RT(t)$: 伸長倍率		
W	: パラメータの重み(0~1)		
$Pi(t)$: ピッチ	$PiAV$: 特定区間内の平均ピッチ
$Po(t)$: パワー	$PoAV$: 特定区間内の平均パワー

高速話速変換の外国語対応(言語の自動判別)

- 音響的特徴の傾向を基に、日本語か英語かを判別するアルゴリズムを検討した
 - (1) 日本語は英語に比べ、各フレーズの途中の短い無音区間が多い
 - (2) 英語は日本語に比べ、語尾(無音区間の直前)の有声度が高い
 - (3) 上記各無音区間の直前50msの有声度を5ms毎に算出し、フレーズ全体で平均化しVとする
 - (4) フレーズ毎に以下の規則で日/英を判定
 $N \geq 0.8 \rightarrow$ 日本語 $N \leq 0.6 \rightarrow$ 英語
0.8 > N > 0.6 の場合、N/Vが0.85以上は日本語、未満は英語
- テレビ番組「ニュースで英会話」の中の10分間の音声を使い、アルゴリズムの優位性を確認した
- 判定結果
10分間のうち、日/英の各フレーズの合計時間長を分母とし、日/英それぞれ正しく判定されたフレーズの時間長の合計を分子とし、正解率を計算
日本語 368.680(秒) / 394.195(秒) = 93.5(%)
英語 78.980(秒) / 114.865(秒) = 68.8(%)

評価実験1

- 日本語の高速音声の聴き取りによる評価実験
- 「横浜市視覚障害者協会」に所属する視覚障がい者20名が被験者
- 日本語能力試験の1級と2級の聴解試験問題から無作為に抽出したものを利用
- 再生速度倍率は、2.0倍速～4.0倍速
- 問題音声の提示から解答を選択し正解率を集計

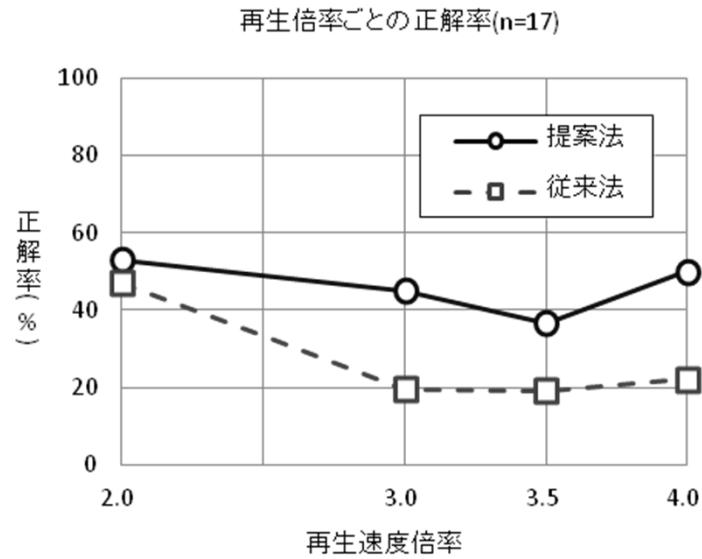


ワイヤレス回答器

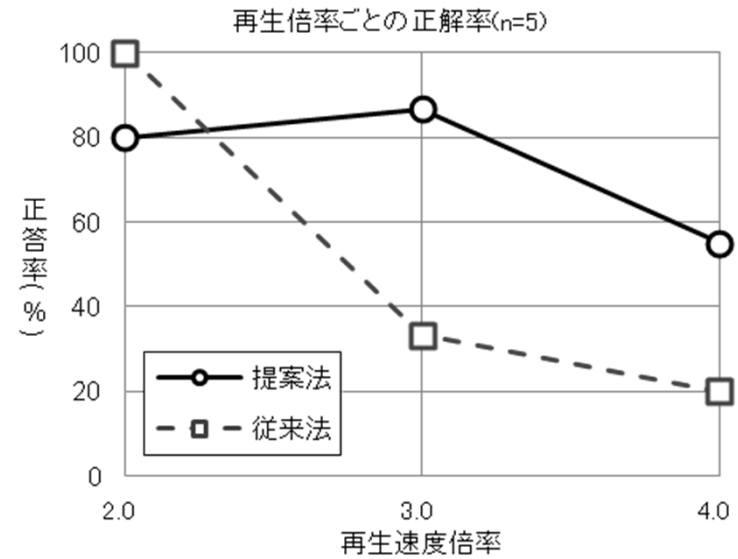


回答風景

実験結果1



通常の被験者群の結果



日常的に高速音声聴取している被験者群の結果

評価実験2

- 英語およびドイツ語の高速音声の聴き取りによる評価実験
- それぞれの言語を母国語とする視覚障がい者が被験者
- TOEICや独検の聴解試験問題から無作為に抽出した音声を利用
- 再生速度倍率は、2.0倍速～4.0倍速
- 従来技術と提案技術の2種類の提示音声から「聴きやすさ」を挙手で回答

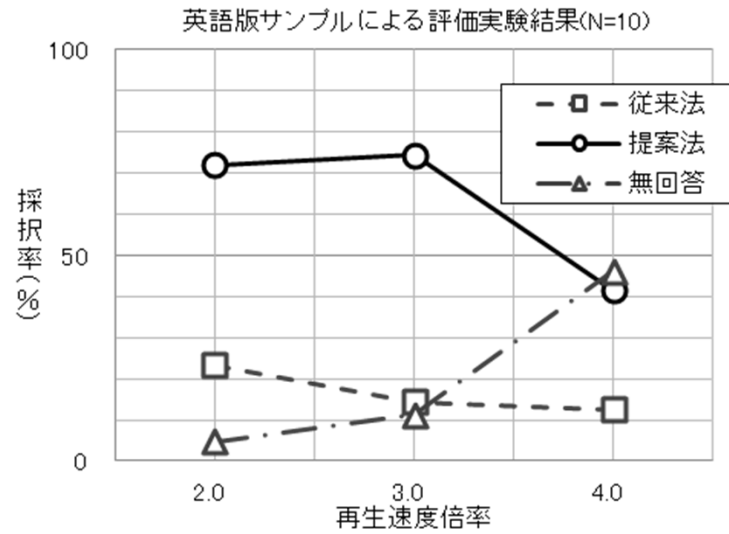


英語実験風景
(カリフォルニア州 サンディエゴ BCCSD)
※BCCSD: Blind Community Center
Of San Diego

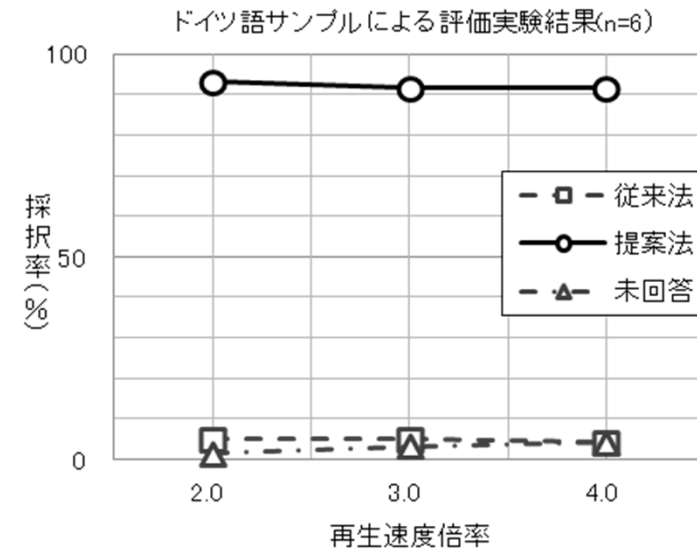


ドイツ語実験風景
(オーストリア ウィーン OBSV)
※OBSV: Austrian Federation of the
Blind and Partially Sighted

実験結果2



英語音声による実験結果



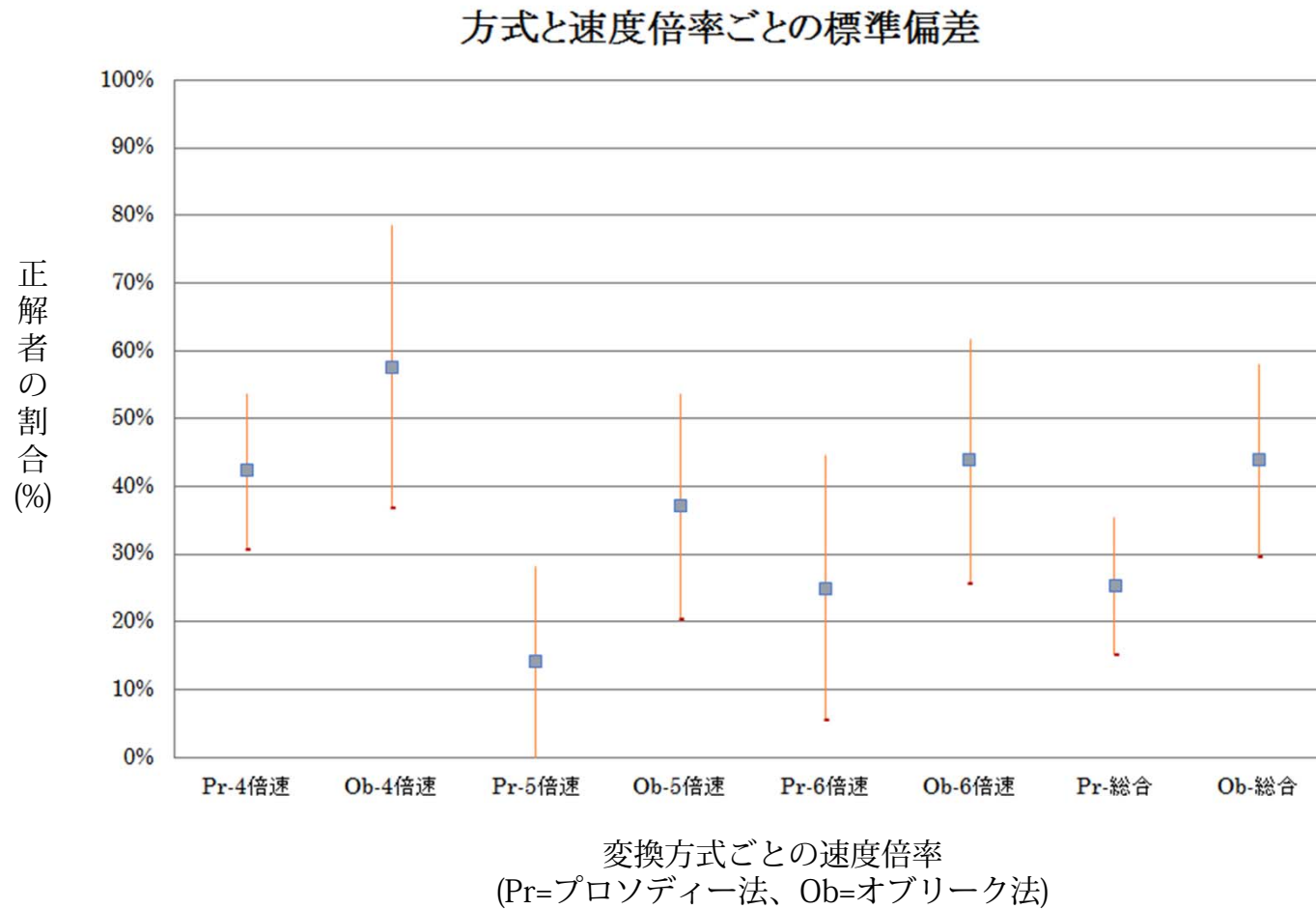
ドイツ語音声による実験結果

評価実験3

- プロソディー法とオブリーク法による「より高速」な音声の聴き取り
- 大和市視覚障害者協会に所属する10名が被験者
- DAISY録音図書(文芸春秋2013年2月号)から抽出した音声を利用
- 再生速度倍率は、4.0倍速～6.0倍速
- プロソディー法とオブリーク法の2種類の提示音声から、内容についての問題を提示し、その正解率で比較

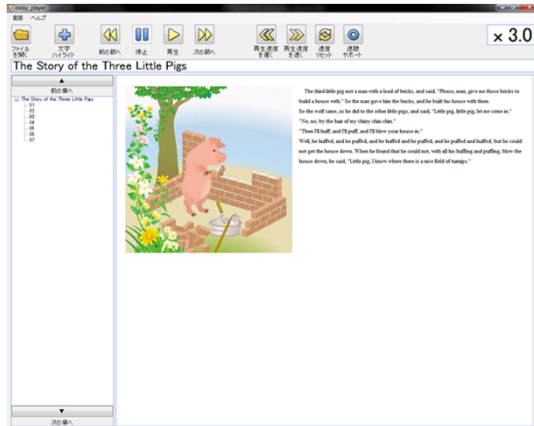
実験結果3

- プロソディー法とオブリーク法による高速音声聴き取りの比較



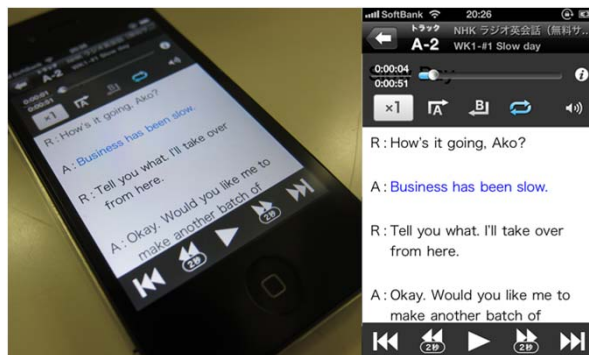
高速話速変換アルゴリズムの実装

【DAISY録音図書再生ソフトウェア】



Windows向けに作成した，DAISY録音図書再生ソフトウェア。DAISY2.02規格に対応しており，0.5～5.0倍速の間で可変速再生が可能。視覚障がい者の利用を想定しており，ガイド音声機能を備え，すべての動作をショートカットキーで実行することができる。リリース予定等については未定。

【語学プレーヤー】



iPhone用の語学練習用ソフトウェア，0.5～3.0倍速の間で可変速再生が出来るため，聴き手の習熟度に応じて自由な速度で聴き取りが出来る。Podcastやオーディオブックコンテンツの再生にも対応し，幅広い利用が期待できる。(NHK出版様より配布)

まとめ

- 高速～超高速度域に対応した高速再生アルゴリズムの開発により、聴き手のニーズに答える効率的な音声情報取得が期待出来る。
- テキスト情報を活用した高品質な高速話速変換が可能になれば、スクリーンリーダーや、電子図書への対応など活用の幅が広がる。
- ハードウェアへの機能実装はメーカーの開発計画との関係から困難。現状ではiOSやAndroidのソフト開発計画が多く、これらのメーカやアプリ開発者向けの、各OS対応の、より使いやすい話速変換モジュールを開発した。
- 視覚障がい者向けだけでなく、学習用途への活用も進行中。