

研究速報

視覚障害者用スクリーンリーダの速度・ピッチ・性別の設定状況

渡辺 哲也[†] (正員)

A Study on Voice Settings of Screen Readers for Visually-Impaired PC Users

Tetsuya WATANABE[†], Member

[†] 国立特殊教育総合研究所、横須賀市

National Institute of Special Education, Yokosuka-shi, 239-0841 Japan

あらまし 視覚障害者用スクリーンリーダの音声合成部改善の基礎資料とするため、スクリーンリーダ利用者の音声設定状況を調べた。最も多かった回答は、音声の速さは最高速度（朗読音声のおよそ2倍）、ピッチは初期値のまま（中間値）、声の性別は男性とした設定だった。

キーワード 視覚障害者、音声合成器、話速、ピッチ、声の性別

1. まえがき

音声合成システムの性能評価では、利用目的への適合性を検討する必要がある[1]。合成音声の重要な応用分野である視覚障害者支援について考えると、音声のみでできるだけ効率良く情報を入手するという目的がある。実際、視覚障害者を対象に筆者らが実施した調査では、合成音声に対する要望として既存のシステムより速い音声を求める回答があった[2]。このニーズがごく一部の利用者のものか、あるいは視覚障害者全般のものを確認するため、スクリーンリーダの話速設定状況を調査した。併せて、一般的な音声設定項目であるピッチと声の性別についても調査した。これらの結果と、晴眼者を対象とした過去の研究[3], [4]とを比較し、音声合成システムに対するニーズの違いを考察し、今後開発が望まれる技術を検討した。

2. 調査の実施

2.1 対象者とスクリーンリーダ

スクリーンリーダを用いてWindowsパソコンを利用している視覚障害者を対象とした。対象スクリーンリーダは、2002年の調査[5]で利用者数の多かった95Reader（システムソリューションセンターとちぎ）、PC-Talker（高知システム開発）、VDM100W-PC-Talker（アクセス・テクノロジー）の3種類とした。

95Readerは、独自のインターフェースを介して株式会社リコーの音声合成器に音声を出力させる[6], [7]。

PC-TalkerとVDM100W-PC-Talkerは、Microsoft Speech API インタフェースを介してProTALKER 97（日本アイ・ビー・エム）[8]に音声を出力させる。いずれもスクリーンリーダから音声合成器へは、読みテキストと音声設定情報（速さ、ピッチ、性別など）が渡される。スクリーンリーダのユーザが設定できるのはこれらの値である。

2.2 調査方法と調査期間

調査は、聞き取りと電子メールの利用の2通りで実施した。聞き取り調査は、盲学校、点字図書館など、視覚障害者が勤務あるいは在学する施設で、2003年4月から同年11月までの間に随時行った。

電子メールによる調査は、「日本理療科教員連名メーリングリスト」（2003年末現在、登録者数：413人）にてアンケートを回覧し、記入結果を電子メールで筆者まで送付してもらう方法である。こちらの調査は2003年9月に行った。

2.3 調査事項

調査では、回答者の属性（年齢、性別、聴覚障害の有無）とスクリーンリーダ使用歴（年数、音声合成装置/ソフト）、現在のパソコン利用状況（使用時間、目的）、そしてスクリーンリーダの種類と音声の設定状況（速さ、ピッチ、声の性別（声質））について尋ねた。

3. 調査の結果

3.1 回答者

全回答者数は36人（男性30人、女性6人）、そのうち8人がメーリングリストからの回答で、それ以外は聞き取り対象者である。年齢は19歳から63歳の間に分布、30歳代が13人と最も多く、平均年齢は38.4歳であった（1人は年齢不明）。音声によるパソコン利用年数は最小値でも2年、最高値は20年、平均値は11.8年（回答者数：35人）だが、10年以上の利用者が27人と回答者の4分の3を占めた。

36人のうち5人からは2種類のスクリーンリーダについて回答を得た。その結果、PC-Talker及びVDM100W-PC-Talkerに対する回答者数は22人となった。これらは同じ音声合成器を使っているため、一つのグループにまとめて集計し、PC-Talkerと表記する。95Readerについては19人の回答者を得た。

3.2 パソコンの利用場面と時間

主として学業を目的としてパソコンを利用している回答者は2人のみで、それぞれ1日当たり7時間と5~6時間利用していると申告した。他の34人は、主に仕事にパソコンを利用しており、その平均利用時間は

3.3 時間（回答者数：30人）であった。学業、仕事のほかに自宅で私的にパソコンを利用する場合もあり、その平均利用時間は1.7時間（回答者数：21人）であった。全体の平均利用時間は4.4時間（回答者数：33人）であった。

3.3 音声設定状況

3.3.1 速度

PC-Talkerの音声速度は、1（最も遅い）から9（最も速い）の間で1段階刻みで設定が可能、初期値は5である。回答者22人のうち17人が、初期速度の5より速く設定していた。最速の9とした回答者が9人で、最も多かった。以下、8と7が3人、6が2人、初期値の5から変更していない人が5人であった。初期値の5より遅く設定した人はいなかった（図1(a)）。最速の9を利用しており、かつ、スクリーンリーダの音声に関する自由記述欄でこれより速い速度を望むと答えた人が2人いた。

95Readerの音声速度は、1（最も遅い）から9（最も速い）の間で設定が可能、初期値は6である。ここでも、すべての回答者が初期値の6以上の速度に設定しており、中でも最速9と回答した人が14人で最も多かった（図1(b)）。最速の9を利用しており、かつ、自由記述欄でこれより速い速度を望むと答えた人が3人いた。

3.3.2 ピッチ

PC-Talkerのピッチ設定は、1（最も低い）から9（最も高い）の間で設定が可能、初期値は4である。ピッチは初期値のままとした人が最も多く22人中13人いた。初期値より高くした人は6人、逆に低くした人は3人であった（図2(a)）。

95Readerのピッチ設定は、1（最も低い）から9（最も高い）の間で設定が可能、初期値は5である。PC-Talker同様、初期値のままとした人が最も多く19人中14人であった。初期値より高くした人は3人、逆に低くした人は2人であった（図2(b)）。

3.3.3 声の性別

PC-Talkerの声質は、標準ではProTALKERの読み秀君（男性の声）と読み子ちゃん（女性の声）のいずれかを選択可能であり、初期設定は読み秀君である。回答者22人のうち19人が読み秀君を選択しており、読み子ちゃんを選択したのは3人であった。

95Readerの声質は、男声、女声、ロボットの3種類から選択可能、初期設定は男声である。回答者19人のうち16人が男声を選択しており、女声を選択し

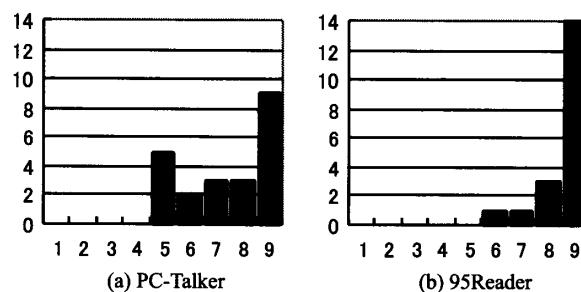


図1 音声速度の設定値。横軸は設定値、縦軸は人数。
Fig. 1 Histogram of setting parameters of speech rate.

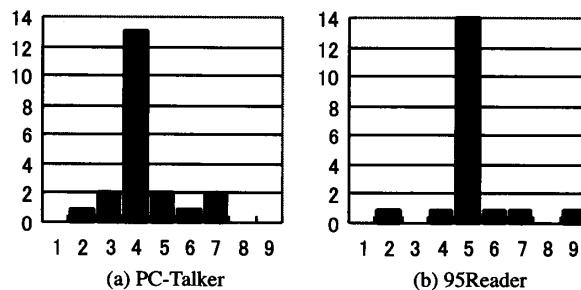


図2 ピッチの設定値。横軸は設定値、縦軸は人数。
Fig. 2 Histogram of setting parameters of pitch.

たのは3人であった。

4. 音声の実測

4.1 速度

PC-Talker XPと95Reader version 5.0の話速を、設定値を変えて実測した。測定用パソコンは、NEC VALUESTAR VL500/2 (CPU: AMD Athlon Processor 946 MHz, メモリ: 240 MByte, OS: Microsoft Windows XP Home Edition) である。読上げ用テキストには、「音声合成システム性能評価方法のガイドライン」(日本電子工業振興協会) [9] の附属書E(参考)総合評価用文セットE.6解説文2の最初の1段落を用いた。その文字数は、漢字仮名交じり文で160字(読点10個と句点3個を含めて計数)、200モーラである。ピッチは初期値とした。PC-Talkerでは「なめらか読みの設定」で「句点を読む」のチェックを外した。

実測の結果を図3に示す。初期値はいずれも400[字/分]前後で、一般的な朗読音声の速度とされる範囲内[10]にあった。最高速は660[字/分]前後となり、一般的な朗読音声の2倍近い速さであった。

パソコンの性能差やOSの違い等による話速の変化の有無を見るため、旧モデルのパソコン(Sony VAIO PCG-N505SR, CPU: モバイル Pentium II Processor 400 MHz, メモリ: 128 MByte, OS: Microsoft Windows 98 Second Edition)に、旧バージョン

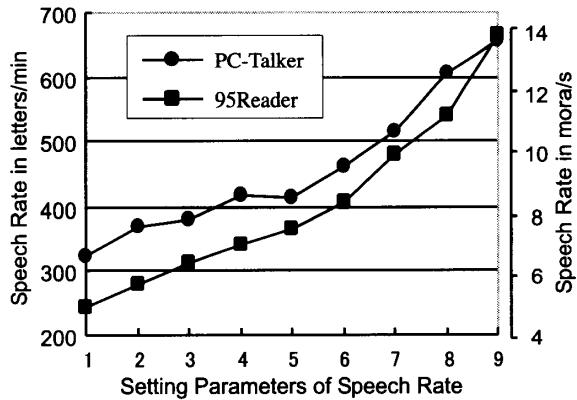


図 3 スクリーンリーダ 2 種の読み上げ速度の実測値（3 回測ったときの平均値）。平均値に対する偏差が最大で 0.5%と小さいため、誤差範囲は記していない。

Fig. 3 Measured speech rate of two screen readers.

ンのスクリーンリーダ (PC-Talker 5 と 95Reader version 3.5) をインストールして読み上げ時間を計測したこと、VL500/2 との違いは、計測点 18 点中 16 点で 1%未満、最大でも 1.7%にとどまった。計測に使った両環境ともスクリーンリーダの動作環境を満たしており、その範囲内ではパソコンと OS の影響は見られないものと推察される。

4.2 ピッヂ

調査結果では、ピッヂを初期値のままとしていた回答者が最も多かったので、両スクリーンリーダの初期値におけるピッヂを実測した。読み上げ単語として、最頻アクセント型である 4 型 (LHHH) をもつ 4 モーラの単語を単語了解度試験用リスト [11] から選び、これらを上記の話速測定条件 (NEC VL500/2) で発声させたとき、両スクリーンリーダとも 4 型で発声し、かつ全モーラとも有声化された単語 5 個を最終的に利用した。読み上げ速度は初期値とした。ピッヂは、SUGI SpeechAnalyzer (アニモ) の「ピッヂ曲線」機能を使って測った。図 4 に示されるとおり、いずれのピッヂも 1 モーラ目で低く、2 モーラ目に高くなり、以後 4 モーラ目まで連続的に下がる。特徴点として、1 モーラ目における最低値、2 モーラ目における最高値、4 モーラ目における最低値を五つの単語について平均をとると、PC-Talker ではそれぞれ 134 Hz, 199 Hz, 163 Hz, 95Reader では 114 Hz, 144 Hz, 117 Hz であった。

5. 考 察

5.1 速度について

晴眼者を対象とした数字の校正作業実験では、被験者の多くが最適とした合成音声の速度は 6.8 と 7.8

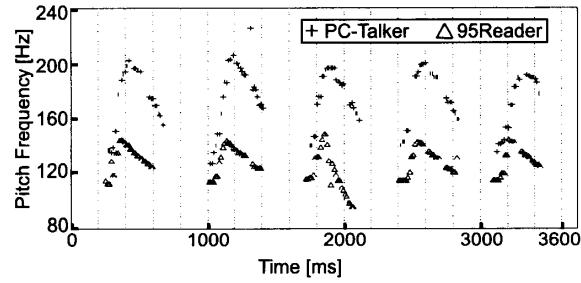


図 4 スクリーンリーダ 2 種のピッヂの実測値。発声させた単語は左から、アマグモ、オヤモト、ソラミミ、タナバタ、ダンタイ。

Fig. 4 Measured pitch frequency of two screen readers.

モーラ/秒であった [3]。一方、本調査では、スクリーンリーダの音声出力を最速に設定している回答者が最も多く、その速度は 2 種のスクリーンリーダでそれぞれ 13.7 と 13.8 モーラ/秒（読み時間には句読点と文節区切りにおける間を含むので、実際の読み速度はこれより速いと考えられる。以下同じ）と、約 2 倍速い。両者の違いは、合成音声の聴取経験の差によると考えられる。校正実験の被験者が合成音声を聞いたことのない初心者であったのに対し、本調査の回答者は数年間の聴取経験を有していた。視覚障害者を対象とした浅川らの実験 [12] でも、「上級ユーザ」2 人の「最適速度」（聴取内容をほぼ 100% 発話再生できた速度）はそれぞれ 1.6 と 2.0 倍速（13.8 と 17.2 モーラ/秒）と、校正実験における最適速度の 2 倍以上であった。

経験による最適速度の上昇を調べるために、音声パソコン利用歴と設定速度との相関係数を計算したところ、PC-Talker で 0.376, 95Reader で 0.077 となり、統計的には両者の間に相関は認められなかった。本調査では、利用歴が最も短い回答者でも 2 年の経験を有しており、この間に適応が済んでいると考えられる。実際、上述の数字校正実験でも、作業経験を積むと速い速度を最適とする割合が増えることが 6 日間の実験期間内に観察されている [3]。

5.2 ピッヂについて

文章の校正作業に適した合成音声のピッヂを求めた実験では、男声では 110 Hz から 130 Hz が適当とされたと報告されている [4]。図 4 を見ると、95Reader はおよそこの範囲に収まっているが、PC-Talker の高アクセント部のピッヂはこれより高く、スクリーンリーダ間で一定の傾向は見られない。このため、視覚障害者と晴眼者の間で聴取に適したピッヂに違いがあるかどうかをここで論じるのは難しい。

5.3 声の性別について

8割強の回答者が男声を使用していた状況は、従来の報告[4]と一致していた。

6. むすび

視覚障害者によるスクリーンリーダの音声設定状況を調べたところ、音声の速さを最高速度（一般的な朗読音声の約2倍）に設定している回答者が最も多く、かつこれよりも速い速度を望む回答者も5人いた。したがって、視覚障害者のパソコン利用支援という分野では、既存の音声合成システムの最高速度よりも速い音声を出力する改良ニーズがあるといえる。その実現のためには、人がどの速度まで正確に聞き取り可能かを心理物理実験等で調べる必要がある。

謝辞 メーリングリストへの調査用紙の配布と、愛媛県立松山盲学校における聞き取り調査を調整して頂いた同校の氏間和仁氏に深く感謝致します。佐賀県立盲学校において聞き取り調査を実施して頂いた同校の中村正徳氏にも感謝致します。両名以外にも調査の実施に御協力頂いた方々、そして本調査に御回答頂いた方々に御礼申し上げます。

本研究は科学研究費補助金（若手研究(B)(2) 課題番号：14780340）により実施した。

文 献

- [1] 素谷英樹，“合成音声技術の評価”，音響誌，vol.48, no.1, pp.46–51, 1992.

- [2] 渡辺哲也、坂尻正次、指田忠司、岡田伸一，“視覚障害者のWindowsパソコン利用状況調査”，信学技報, SP2000-47, 2000.
- [3] 素谷英樹、糟谷清基、森田清紀，“校正支援用規則合成音声の発声速度の役割”，音響誌, vol.47, no.2, pp.96–98, 1991.
- [4] 素谷英樹、田島一幸，“合成音声の聴き易さとピッチ特性の関係”，信学技報, SP88-153, 1989.
- [5] 渡辺哲也、指田忠司、長岡英司、岡田伸一，“視覚障害者のWindowsパソコン及びインターネット利用・学習状況”，信学技報, WIT2002-62, 2002.
- [6] 渡辺哲也、岡田伸一、伊福部達，“GUIに対応した視覚障害者用スクリーンリーダの設計”，信学論(D-II), vol.J81-D-II, no.1, pp.137–145, Jan. 1998.
- [7] リコーおもしろ科学館, <http://www.ricoh.co.jp/omoshiro/voice/voice4.html>
- [8] ProTALKER 97, <http://www-6.ibm.com/jp/voiceland/pt20/>
- [9] 音声合成システム性能評価方法のガイドライン, 日本電子工業振興協会, 東京, 2000.
- [10] 三留幸夫，“合成音声の応用と今後の展望”，音響誌, vol.49, no.12, pp.875–880, 1993.
- [11] 坂本修一、鈴木陽一、天野成昭、小澤賢司、近藤公久、曾根敏夫，“親密度と音韻バランスを考慮した単語了解度試験用リストの構築”，音響誌, vol.54, no.12, pp.842–849, 1998.
- [12] 浅川智恵子、高木啓伸、井野秀一、伊福部達，“視覚障害者のスクリーンリーディングにおける最適・最高速度”，ヒューマンインターフェースシンポジウム 2002, pp.543–546, 2002.
(平成16年12月15日受付, 17年3月14日再受付)