

観光情報システムの音声合成導入に関する考察

河本祐幣[†] 市川尚[†] 窪田諭[†] 狩野徹[‡] 阿部昭博[†]

岩手県立大学大学院ソフトウェア情報学研究科[†] 岩手県立大学社会福祉学部[‡]

1 はじめに

筆者らの研究グループは、携帯電話を用いた観光情報システムの開発を行ってきた。観光情報システムとはBluetoothタグ等を用いて、観光スポットに応じた情報提供を行うシステムである。これまでの研究から、音声での自動案内が有用であることがわかっている¹⁾が、観光情報システムはユーザー特性（視覚障害者、子供など）ごとにコンテンツがあるため、音声コンテンツの作成に多大な時間を要していた。また、外国人観光客の増加によりコンテンツの多言語化が求められている。その他にも音声での情報提供に関して、観光客から音声案内の速度や話者変更の要望がある。これらを踏まえると観光情報システムにおける音声提供に関して、音声合成の導入を行うことが現実的であると言える。

ものづくりや情報システム導入を行う際の重要な3つの要素としてQuality（品質）、Cost（価格）、Delivery（納期/時間）がある。本研究では音声コンテンツ作成方法をQCDの観点から分析し、観光情報システムの音声合成の導入に関して考察する。

2 QCDからみた音声コンテンツ作成方法の特徴

2.1 音声コンテンツ作成方法の分類

観光情報システムにおいて音声で情報を提供するには、事前に音声コンテンツを作成する必要がある。音声コンテンツの作成方法として、自身で録音等を行う場合（内製）と外部へ委託する場合（外注）、音声合成を活用する場合の3通りが挙げられる。なお、音声合成ソフトはサーバ導入型とデスクトップ上で動作するデスクトップ型の2種類に分けられる。また今回想定する音声合成導入組織は、情報システム環境や運用体制が既に整備されている自治体や観光施設である。

2.2 品質

音声合成による案内の際の重要な評価指標として合成音声の明瞭性（聞き取りやすさ）がある²⁾。

内製：ナレーションが不慣れであり、録音の際にノイズ等が入るため、明瞭性はあまり高くない。

外注：プロによるナレーションのため明瞭性は高いと言える。

音声合成：音声合成ソフトではパラメータ（話者、

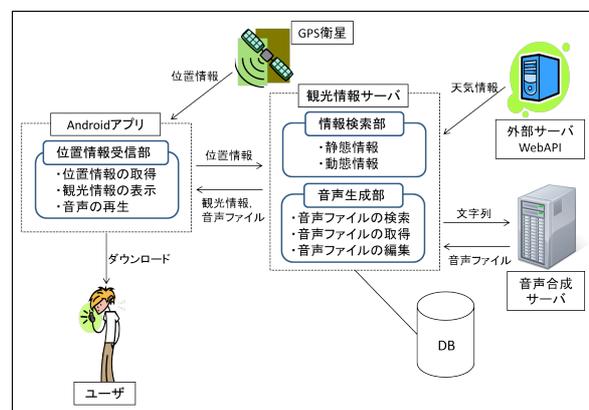


図 1. システム構成図

速度、ピッチ等)の変更が行える。特にサーバ導入型では動的作成も可能であるが、合成音声の明瞭性に関しては把握していないため、検証の必要がある。

2.3 コスト

内製：通常業務外のため、多少の人件費がかかる。
外注：1コンテンツごとにコストが発生する。

音声合成：インシヤルコストとしてソフト導入費が、ランニングコストとして年間ライセンスなどがかかる。なお、必ずしも導入費とライセンス、双方のコストがかかるわけではない。またデスクトップ型では、動的な音声ファイルの作成が行えないため、ソフトの価格が低く設定されている。

2.4 納期（時間）

内製：コンテンツの読み上げ練習・録音・変換作業により、作成に時間がかかる。

外注：業者とのやりとりなどにより、発注から受け取りまでに時間がかかる。

音声合成：音声コンテンツの作成自体に時間はかからないが、サーバ導入型では音声コンテンツ作成環境の構築に時間がかかる。またデスクトップ型ではソフトの使用方法を理解する必要がある。

3 音声合成による音声コンテンツの品質検証

3.1 システム概要

合成音声の明瞭性や話者・速度変更の有用性検証のため、音声合成を使用した観光情報の提供を行うシステムの開発・運用を行った。本システムの構成を図1に示す。静的情報とは時間的経過等によりあまり変化しない情報（スポット情報等）であり、動的情報とは時間的経過によりその内容が変動する情報（イベント情報等）を指す¹⁾。本システムの特徴を以下に示す。

A Study on Introduction of Speech Synthesis for Tourist Information System
Yuhei Kawamoto[†], Hisashi Ichikawa[†], Satoshi Kubota[†], Toru Kano[†], Akihiro Abe[†]

[†] Graduate School of Software and Information Science, Iwate Prefectural University

[‡] Faculty of Social Welfare, Iwate Prefectural University

- 1) 音声合成を使用した情報提供：音声合成を活用した音声コンテンツの作成及びその提供を行う。また、音声速度（早い、普通、遅い）や話者（男性、女性）の変更を行える機能を実装することにより、聞き取りやすい音質を選択可能にする。なおデフォルトの音声は速度が普通、話者は男性とした。
- 2) プッシュ型の情報提供：GPS による位置の同定及びそれに応じた観光情報の提供を自動的に行うことでユーザビリティに配慮する。

今回は、パラメータ変更などが可能な株式会社AITalk のサーバ導入型音声合成ソフトを選定した。

3.2 評価

10月20日から2ヶ所の観光地で本システムの運用を行っている。また、2011年11月7日に平泉毛越寺にて被験者6名（60～70代：5名，40代：1名）を対象に評価実験を行った。Android アプリをダウンロードした端末を3名1グループに1台貸し出し、実際にシステムを利用しながら毛越寺を散策してもらった後、グループインタビューを行った。主な質問項目は合成音声の明瞭性、速度変更の有用性、話者変更の有用性の3つである。

3.3 結果

運用の結果、12月22日までで53人のユーザが使用している。そのうち速度変更を行った人数は13人（約25%）、話者変更は17人（約32%）であり、速度・話者変更のニーズがあることがわかった。

グループインタビューにおいて、合成音声の明瞭性に関しては、全員が問題なく聞き取れたと回答した。速度変更に関しては、通常で十分聞き取れるという意見が多かったが、景観を見ながらだと遅い速度の方が良いという意見も挙げられた。話者変更に関しては、男性の方が聞き取りやすい、女性は音程が高いので聞き取りにくいという回答を得た。

4 考察

4.1 品質

合成音声の明瞭性に関しては、全員から肯定的な意見が挙げられ、機械の声でも十分に案内が可能であると言える。また速度変更の有用性も示され、ユーザが自身の環境に合わせて速度を変えることで、更に明瞭性が上がると考えられる。話者変更では、男性の声の方が聞き取りやすいという意見が多数挙げられたが、今回選定した女性の音程が比較的高めであるからと考えられる。女性の声で案内を行う場合は、音程が低めの方が好ましいことが示唆された。一方で音声合成以外の作成方法では、速度・話者ごとに音声コンテンツを作成する必要があるため、コストまたは納期の負担が増加すると考えられる。

4.2 コスト

音声合成（サーバ導入型）を使用した場合と外部へ委託する場合のコストの比較を行った結果、外部

表 1. 音声ファイル作成方法とその特徴

		内製	外注	音声合成	
				デスクトップ型	サーバ導入型
品質	明瞭性	低～中	高	低～中	低～高
	パラメータ可変性	低	低	中～高	中～高
	動的作成	不可	不可	不可	可
コスト	対応可能な情報の分類	静態	静態	静態	動態、静態
	イニシャル	-	-	数万円	数十万円
	ランニング	-	-	中	中
	ファイル作成	低	高	-	-
	納期（時間）	高	高	低	低

委託において30秒（約200文字）程度の音声コンテンツ約40個の依頼でイニシャルコストが償却できることがわかった。今回は静的な情報（スポット情報等）を対象としていたが、動態情報を含めると導入効果は更に向上すると考えられる。

4.3 納期

音声合成を使用した場合と自身で作成した場合の音声コンテンツ作成時間の分析を行った。以前は学生が音声コンテンツの作成を行っており、1つの音声コンテンツにつき全体で30分程度かかっていた作業が音声合成を導入した結果、5分程度に短縮された。今回の分析時には既に音声合成を活用した音声コンテンツ作成環境が構築済みであり、音声合成（サーバ導入型）を導入した場合には、環境構築に時間を要するが、構築にかかる作業は一度であり、長期的に見ると音声合成を導入した場合の方が作成時間は大幅に短縮されると考えられる。

4.4 結論

3つの音声コンテンツ作成方法（内製、外注、音声合成）に関してQCDの観点から分析を行った結果を表1にまとめる。観光情報システムにおいては、提供する音声コンテンツの数が多く、動態情報提供の必要もあるため、分析の結果から、サーバ導入型の音声合成ソフトの導入が望ましいと言える。

5 おわりに

今回は静態情報のみの提供であったが、イベント情報などの動態情報の音声提供の必要もある。今後は動態情報も音声合成を用いて提供することで、観光情報システムにおける音声合成導入の有用性を検証していく予定である。

参考文献

- 1) 河本祐幣他：音声合成を利用したUD観光情報システムの拡張，情報処理学会第73回全国大会，3ZE-2(2011)。
- 2) 三井康行：音声合成の活用シーンと要求される品質との関係，情報処理学会研究報告，SLP-83-8(2010)。