

# Usability assessment of text-to-speech synthesis for additional detail in an automated telephone banking system

Hazel Morton<sup>a</sup>, Nancie Gunson<sup>a,\*</sup>, Diarmid Marshall<sup>a</sup>, Fergus McInnes<sup>a</sup>,  
Andrea Ayres<sup>b</sup>, Mervyn Jack<sup>a</sup>

<sup>a</sup> *Centre for Communication Interface Research, The University of Edinburgh, Alexander Graham Bell Building,  
King's Buildings, Edinburgh, EH9 3JL, UK*

<sup>b</sup> *Lloyds Banking Group, Canons House, Canons Way, Bristol, BS99 7LB, UK*

Received 4 February 2010; received in revised form 29 April 2010; accepted 31 May 2010

Available online 31 July 2010

---

## Abstract

This paper describes a comprehensive usability evaluation of an automated telephone banking system which employs text-to-speech (TTS) synthesis in offering additional detail on customers' account transactions. The paper describes a series of four experiments in which TTS was employed to offer an extra level of detail to recent transactions listings within an established banking service which otherwise uses recorded speech from a professional recording artist. Results from the experiments show that participants welcome the added value of TTS in being able to provide additional detail on their account transactions, but that TTS should be used minimally in the service.

© 2010 Elsevier Ltd. All rights reserved.

*Keywords:* Text-to-speech (TTS); Usability; Dialogue design; Automated telephony

---

## 1. Introduction

Speech applications have two primary options for speech output: natural speech prompts, recorded from human voice actors, and synthesised speech. Many early uses of synthesised speech, or text-to-speech (TTS) synthesis, were in systems for accessibility, for example reading systems for blind or sight impaired computer users, and mainstream usage of TTS was “severely limited by its quality” (Taylor, 2009: p. 2). However as the quality of TTS systems improves, where quality defined in terms of the intelligibility of the system and the naturalness of the voice, TTS becomes more common in everyday applications.

In the creation of speech systems there is a trade-off to be made between the quality and expense of recorded prompts against the flexibility of synthesised prompts. Recorded prompts, which although having the benefit of sounding natural, can be expensive to create as they require the recording time of a voice actor. Synthesised speech may sound less natural, but has the advantage of being more flexible as the service designer can create new prompts as and when required without having to visit the recording studio. The use of TTS could be particularly beneficial therefore in services that require to output dynamic information, such as place names or company names, where recording such a diverse set

---

\* Corresponding author. Tel.: +44 131 651 7120; fax: +44 131 650 2784.  
E-mail address: [Nancie.Gunson@ccir.ed.ac.uk](mailto:Nancie.Gunson@ccir.ed.ac.uk) (N. Gunson).

of prompts would be unfeasible. Importantly, the use of TTS in such a case can potentially add value to a system that otherwise would be more limited in the information it can provide.

Previous research has investigated the usability and effectiveness of synthesised speech in a variety of applications, for example in a flight information system (McInnes et al., 1999), in a personal information management application (Gong and Lai, 2003), in tutoring applications (Baylor et al., 2003; Forbes-Riley et al., 2006) and in a smart-home system (Möller et al., 2006).

Research which investigated users' perceptions of the personality of a synthesised voice compared with a recorded voice (on which the TTS system was modelled) found that the synthesised voice is associated with more negative personality characteristics than the recorded voice (Love et al., 2000). However other research which investigated synthesised speech in comparison to a number of recorded speech samples in a smart-home system found that synthesised speech prompts do not necessarily receive more negative ratings than recorded speech (Möller et al., 2006). Further, investigation was made of a combined recorded and synthesised voice compared to a fully synthesised voice. It was found that the combined recorded and synthesised version scored significantly higher than the fully synthesised version on overall quality, voice adequacy and voice pleasantness. However, no significant differences were found for listening effort. This study recommends that, as much as possible, recorded voices should be used and supplemented with synthesised when required, rather than opting for a fully synthesised system.

In the evaluation of TTS systems, many empirical evaluations focus on the acceptability, naturalness and comprehensibility of the systems (Stern et al., 1999; Stevens et al., 2005; Viswanathan and Viswanathan, 2005). Such research focuses on the comprehension or acceptability of TTS as a speech solution, that is, assessing TTS system prompts solely from a quality perspective. However, even if it can be assumed that the quality of TTS speech prompts are not as good as recorded prompts, the use of TTS in a dialogue system can be beneficial to its users by providing additional information that would be not be viable as a recorded prompt solution. Thus it is important to evaluate the use of TTS as a speech output solution from a usability perspective, within the context of a real-world application. The four studies described in this paper detail the evaluation of the usability of TTS within an already established dialogue system.

## 2. Evaluation of TTS: four studies

Four studies are presented here which investigate the use of TTS in an automated telephone banking service from a usability perspective. The service used in the experiments is a mirror copy of a telephone banking service from a major UK bank, referred to here as the Case Bank. The existing service at the time of the research utilised service prompts recorded from a human voice-talent actor (a female Southern British English voice) and the system functionality allows users to access their bank accounts, find out balance and transaction information and complete simple banking tasks such as transferring money between their accounts. The system utilises a speaker independent commercial speech recogniser<sup>1</sup> so that users can interact with the system using speech input without any prior training of the system; dual tone multifrequency (DTMF) input is also available. Natural language understanding (NLU) is implemented in the system via a finite-state grammar, in which allowable sequences of words and phrases are hand-coded. Each path in the grammar is associated with an appropriate feature-value pair in order to extract the meaning of the utterance.

Each of the experiments described here focused on the transaction listings within the banking service which provides a number of pieces of information on recent banking transactions such as the date and the amount of the transaction (e.g. "on the 12<sup>th</sup> of January, a debit for £30"). These prompts are concatenated from a library of recorded speech prompts. However, the system was unable to provide more detailed information, such as the exact location of a cash withdrawal transaction or the retailer to which a debit card payment was made. In employing TTS in the system development, such information could be used and the relevant speech output created dynamically in order that the information be passed on to the user (e.g. "on the 12<sup>th</sup> of January, a debit for £30 to *The Gift Shop*").

This series of experiments investigates customer perceptions of the use of TTS in the automated telephone banking system for this purpose. The first experiment investigates the usability of both a fully and a partially synthesised system for recent transaction readouts in comparison to a fully recorded system. The second experiment investigates the use of minimal TTS (for less frequently occurring names only) in providing additional detail. The third experiment investigates the optimal location of the TTS component within a recorded speech prompt. The fourth experiment investigates the

<sup>1</sup> Nuance v8.0.0 [www.nuance.com](http://www.nuance.com).

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات