

ウェブビジネスにおける音声読上げ技術 の可能性 food.com を超えて

鈴木 淳弘

1. はじめに
2. food.com
3. TTSアプリケーションプロバイダー
4. 音声読上げ技術、音声認識技術と自動翻訳技術
5. 日本参入の可能性
6. food.comの課題
7. 結び

要旨

本稿では先駆的な会社、food.com を例にとって、米国における音声読上げ技術（TTS: Text to Speech）の利用を紹介する。すなわちTTSによって、ウェブビジネスを展開する企業は比較優位を得ることができる。TTSを自社のウェブサーバーに導入すると、顧客に24時間サービスを容易に提供できるようになる。TTSにより人手のかからない、つまりコストがかぎりなく低いサービスを実現できる。また、TTSと密接不可分の技術である音声認識技術（STT: Speech to Text）と自動翻訳技術（MT: Machine Translation）の結びつきにより、さらに強力なビジネスツールが生み出されることだろう。最後に、food.comは、今よりも多くのインテリジェンスを持つ無数の端末に特徴づけられる将来の広帯域自律分散型のネットワーク（サービス）への過渡期に現れたことの意義も指摘する。

1. はじめに

音声読上げ技術（TTS: Text to Speech）を利用することによって、ウェブビジネスを展開する企業は比較優位を得ることができる。TTSを自社のウェブサーバーに導入すると、顧客に24時間サービスを容易に提供できるようになる。TTSにより人手のかからない、つまりコストがかぎりなく低いサービスを実現できる。加えて、TTSは人手を介しない分、人為ミスの可能性を皆無にしつつ、正確な業務をこなすことができる。

例えば、ユーザーは特定のボイスメールに電話をかけて、暗証番号などをダイヤルすれば、ファックスや電子メールを外出先からでも、普通の電話で聞くことができる。また、銀行の場合、顧客からの残高照会等の問い合わせには、マシンによる自動応答をしていることはよく知られている¹。しかし、こういった単純なTTSの特徴を生かして、インターネットでビジネスをしている企業は意外と少ない。本稿では先駆的な会社、food.comを例にとって、米国におけるTTSの利用を紹介する。そして、TTSと密接不可分の技術にも目を向けて、それらとインターネットを結び合わせることによるビジネスの可能性を考える²。

2. food.com

food.comは、TTSを利用して、ユニークなウェブサイトビジネスを展開している。顧客はfood.comのウェブ上で電子メールで注文すると、food.comのサーバー上のTTSにより注文が音声に変換されて、注文先のレストランに電話がかかり、その注文が読み上げられる。レストランが注文を受けると、顧客に注文受け取り確認メールをサーバーが配信するのである。

food.comは、1996年11月にCybersliceとして、ティム・グラス（Tim Glass）によって設立された。彼は、映画「The Net」で、主演のサンドラ・ブロックがパソコンでピザを注文するシーンをみて、このビジネスを思い付いた。この映画はインターネットの専門家からは少々非現実的と相手にされなかったが、グラスは直ちにこのおとぎばなしを実現させた。food.comは食事のデリバリーサービスを希望する顧客の住所にもとづいた様々なレストランを紹介して、注文を受け付けるサービスをインターネット上で全米で初めて展開した。グラスによれば、2005年には、だれでも家からテレビの「食事チャンネル」で注文できるようになるし、アメフトの試合をみていると、画面に「ここをクリックすれば、ピザを届けます」というメッセージが流れるようになるという。そうなったら、電話でわざわざ食事を注文することもなくなると考えている。food.comは設立後1年も満たないうちに5000軒のレストランと契約をするにいった³。

顧客は実際どのように注文をするのだろうか。まず、food.comのホームページにアクセスする。そして、デリバリーかテイクアウトを選んで、自分の住所を入力する。すると、近辺のレストランの一覧が表示される。その中の一軒を選ぶと

¹ 例えば、シティバンク。0120-110-330

² <http://www.food.com>

³ <http://news.cnet.com/news/0-1007-200-323513.html> Cyberslice から cybermeals.com へと社名変更。

メニューが表示される。支払いはクレジットカードが基本のようだが、現金を認めているところも多い。そして注文を受け付けた food.com のシステムは、テキストベースで受け付けた注文を、TTS技術を利用して、音声に変換する。そして、システムがレストランに電話をかけて、合成された音声で、注文を伝える。テイクアウトを選んだ顧客は、注文をするとレストランの所在を示した地図を得る。これは、MapQuestとパートナーシップを結んでいるためだ⁴。またデリバリーの場合、food.comからレストランには顧客の住所通知と支払が行われる。このシステムは顧客の住所や好みなどの履歴をデータベース化して、次回以降の注文を効率的に処理している。すでに、50万人以上の顧客を登録している⁵。

food.comは、レストランから複数の収入を得ている。レストランは、まず、メニューや位置情報の初期登録に400ドルを支払う。そしてfood.comは、レストランからウェブサイト構築費用として5,000ドルを徴収する。あとは、注文ごとに3~10%の手数料を徴収することになる⁶。

2.1 food.com の比較優位性

TTS技術を利用した注文受付を導入して、food.comはほかのインターネット上のCuisineNet⁷やFood Near You⁸などのレストラン情報提供サービス⁹との差別化を図ったと言えよう。

第一に、TTSによりfood.comは顧客からの注文を受けたり、レストランに注文を伝えたりするための要員を持たずに、24時間サービスを提供できる。2万4000軒のレストランと50万人の顧客とを抱えている現在でも、社員は150人程度に過ぎない。残念ながら food.comは非公開企業であるので、財務状況を入手できないが、オペレーションにかかる人件費は低いと思われる。

第二に、food.comは "internet food delivery" という分野において最初に企業ブランドを打ち立てた。Amazon.comやe-bayなどのウェブビジネスでの成功を収めている企業などから見れば明らかなように、あたらしいビジネスモデルを最初に打ち出しブランドを確立することは強力な比較優位となる。food.comもそのブランド力をもとに、いろいろなメディアとパートナーシップを結び、顧客を増やしている。一例として、AOLとの場合を紹介する。まず(潜在的な)顧客がホームページのトップページにアクセスすると、"Daily Essentials"として"Top News"や"Weather"など14のカテゴリーと、"Web Centers"として"Autos"や"Entertainment"など20程度のカテゴリーが現れる。この"Web Centers"のひとつの"Food & Cooking"を選択すると下のような画面になり、food.com のバナーがトップに出てくる。全米一の1400万人近い会員を擁するAOLのサイトを利用して、food.comの認知度は高められていよう。その他にロサンゼルスタイムズからのリンクもはっている¹⁰。

⁴ <http://www.mapquest.com/>

⁵ <http://www.food.com/cgi-bin/WebObjects/food.woa/01261000001160200000388720000071/Main wo/252510000061/29.0/4/appserv03?350,12>

⁶ 1998年1月現在。food.comへの電話インタビューによる。

⁷ <http://www.dinenet.com/index.shtml>

⁸ <http://food.from.net/>

⁹ これらはレストランの住所と連絡先を提供するのがメインであって、注文を受けることは副次的なものと考えているようだが。

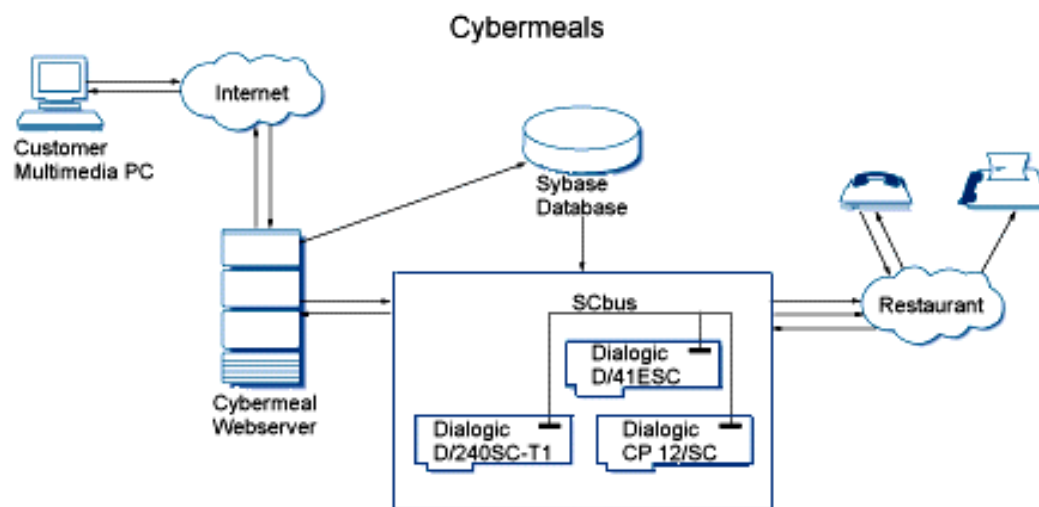
¹⁰ http://www.calendarlive.com/LA_Times/Arts_and_Entertainment/Restaurants_Cafes/



第三に、ブランドの優位と密接に関係するが、レストランと顧客にとっても food.com の「ドミナント性」にメリットがあることは明らかであろう。まず、レストランはインターネットを利用して顧客の拡大を目指す場合、food.com のこの分野におけるシェアが圧倒的に高いのであれば、迷わず food.com をそのパートナーに選ぶであろう。food.com 以外の類似サービスをする会社に二重投資をする手間を省けるのであるから。また、顧客にとっても、food.com がインターネットで注文を受け付けてくれる自宅そばのレストランを網羅しているのであれば、迷わず food.com を選ぶだろう。お腹の空いた顧客が、Yahoo のサイトで「レストラン」を入力して検索したところで、何千軒ものヒットがあって、その99%は自宅から遠くはなれているといった不都合も解消される。この food.com のブランド力とドミナント性が food.com をさらに市場での強さを高めるといふ、好循環を引き起こしている。

2.2 food.com のシステム構成

food.com のシステムは音声読み上げ技術や音声認識技術の分野の最大手のひとつ、ダイアログ社 (Dialogic) によるものである¹¹。



¹¹ <http://www.dialogic.com/coolapps/cyberblock.htm>

ウェブサーバー	CPU: ペンティアムII 200MHz以上 RAM: 128MB以上 HDD: 9GB以上
電話回線収容装置	Dialogic D/41ESC 音声処理ボード Dialogic D/240SC-T1 音声処理及びネットワーク収容ボード Dialogic CP12/SC ファックスボード SCbus プラグインモジュール
データベース	サイベース社の汎用パッケージ
TTSアプリケーション	センチグラム社のTruVoice
food.comのシステム構成	

food.comのシステムは簡易で安価である。システムを構成する主なものとして、ウェブサーバー（1万ドル程度）、データベース¹²（1,000ドル程度）、音声およびファックスボード（1万5,000ドル程度）が挙げられる。その他のパーツ（TTSアプリケーション¹³など）の価格の詳細がわからないのであくまで筆者の推計ではあるが、システムの構築には約3万ドル程度しかかからない。つまり、重要なのは、上記に示したように、このシステムがfood.com向けに特別に作られたのではない。つまり、誰にでも手に入れられる市販品で成り立っていることである。

3. TTSアプリケーション プロバイダー

food.comが採用しているセンチグラム社のほかに、TTSアプリケーションを扱っている主な企業を紹介する。

ラノート・アンド・ホースピー社（Larnout & Hauspie）は、TTSアプリケーション市場の最大手といわれている¹⁴。主力商品のTTS3000は、英語のほか、ドイツ語、フランス語、スペイン語、イタリア語、韓国語と日本語に対応している。マイクロソフト社のOS、Windowsにのみ対応する¹⁵。後述するように、この会社はTTSを足がかりに戦略的に事業の拡大を図っている。

エロクエント社（Eloquent Technology）はTTSアプリケーション分野に特化している¹⁶。主力商品であるETI Eloquentの特徴は、出力音声のパラメーターを変えられることである。男性、女性、子供の声やスピードの異なる声なども自由に設定できる。しかし、テキストにある文法ミスを補正したり、再生されている音声を聞き手がストップしたり、リピートさせることはできない¹⁷。英米語、フランス語、スペイン語、ドイツ語、イタリア語に対応。ウェブサイトでデモンスト

¹² <http://www.sybase.com/products/databaseservers/>

¹³ <http://www.itl.atr.co.jp/comp.speech/Section5/Synth/truvoice.html>

¹⁴ <http://www.lhs.com/>

¹⁵ <http://www.lhsl.com/ssyn/tts3000m.asp>

¹⁶ <http://www.eloq.com/>

¹⁷ <http://www.eloq.com/eti0elo.html>

レーションを行っている¹⁸。

コンパクト社は昨年を買収したデジタル・イクイップメント社 (Digital Equipment) のDECTalkを引き続き取り扱っている¹⁹。ETI Eloquentと同様に、様々な音声を出力できる。英米語、スペイン語とドイツ語に対応。WindowsとUNIX上で作動する²⁰。

上記のほかに、アーティソフト社 (Artisoft)²¹のVisualvoice、ディレクトエクストラ社 (DirectXtra) のXpress Xtra²²などがあげられる。

4. 音声読上げ技術、音声認識技術と自動翻訳技術

まず、TTSの可能性を考えてみる。現在では、電話で電子メールを読み上げるのが、一番多い利用法であろう。しかし、人手を使わずに、24時間サービスを提供できる利点を生かすには、food.comのようにデータベースに貯えた情報を電話で自動的に再生するのがよい。例えば、米国西海岸のあるビール醸造会社では、販売店との取引状況を24時間提供できるようにした。受発注の状況、生産状況、出荷・配送状況などのデータベースにある情報を、顧客がいつでもアクセスできるようにした。顧客は注文したビールが何時に配送される、ということまで把握できるようになった。一方、醸造会社側はオペレーションの人件費が要らなくなり、1万5,000ドル/月のコスト削減を達成した²³。

最近では様々な音声認識 (汎用) ソフトウェアが実用化されている。R&Dの長年の研究テーマであった音声認識 (STT: Speech to Text) 技術は、コンピューターが言葉を聞き、理解して、その結果を文章にすることを目指している。従って、このSTTとTTSを組み合わせれば、電話回線を通じてインターネットと接続しているすべてのコンピューターは、人々と両方向の会話を行うことになる。現時点のSTT技術では、例えば、カーズウェイル・アプライド・インテリジェンス社 (Kurzweil Applied Intelligence) は、有名な財務ソフトのクイッケン社²⁴にコマンドや数値の音声入力を実現している²⁵。STT技術をウェブサーバーに組み込むことにより、ウェブビジネスを行う企業は、顧客が何を欲しているのかを「話して」もらうことができる。

STTと自動翻訳 (MT: Machine Translation) の結びつきにより、さらに強力なツールが生み出されよう。企業と顧客はお互いの言語を意識することなく、コミュニケーションをすることができるのである。MTは、MTエンジンと呼ばれるソフトウェアにより、入力された文章をゼロから翻訳する。仮にSTT技術が完成されたのもであれば、例えば、顧客が口頭で日本語で行った注文を文章化 (日本語) し、MTにより英文に翻訳して、最後にTTS技術によって音声による英語の注文になる。しかし、STT技術により完全な文章が常に作成されることは難しい。しかし、

¹⁸ <http://www.eloq.com/demos.html#Text Processing>

¹⁹ <http://www.digital.com/oem/products/dectalk/dtalk.htm>

²⁰ <http://www.digital.com/oem/products/dectalk/dtproducts.htm>

²¹ <http://www.artisoft.com/>

²² <http://www.directxtras.com/>

²³ <http://www.dialogic.com/company/whitepap/3483web.htm>

²⁴ <http://www.quicken.com/>

²⁵ <http://www.voicerecognition.com/kurzweil/> L&H社に買収された

現時点のTTS技術ではタイプミスや文法のミスのある文章の処理は難しいので、結局のところSTTやMTとの連携は進んでいない。さらに、MTの開発は、英語と自国語の翻訳という英語圏以外の様々な言語の人々のニーズのほうが大きいので、各国でバラバラに進んでしまっている²⁶。当面、日本をはじめ非英語圏の国はSTTやTTSの研究・開発も進んでいる米国との連携が重要であろう。残念ながら、MT、STT、TTSを組み合わせたアプリケーションを、インターネット全盛の今期待するのは無理なようである。

しかし、最近のCTI (Computer Telephony Integration) の発展が、1993年のTAPIと1997年のTSAPIという2種類のプログラミング・インターフェースの標準化を契機としていることは示唆に富む²⁷。これにより、CTIのアプリケーションを、使用する機器のベンダーに係りなく選べるようになったのである。そうだとすれば、各国で開発されている自動翻訳のシステムの中で共通なものを抽出し、モジュール化して、CTIのインターフェースと整合させることはできないであろうか。コンピュータの情報処理機能と電話の通信機能を融合するCTIに、STT、TTSとMTが統合されて実現されるサービスは、基礎となる技術は同じであろう。もし、CTIにMT、TTS、STTが合わされば、例えば、世界規模でバンキングサービスを展開しているCITIグループが、各国においているコールセンターを、究極的には1ヶ所に集約することも可能であろう。

音声ソフトウェアの最大手のL&H社は、TTS、STTとMTの技術の融合をM&Aによって推し進めている。1996年に、まずバークレイ・スピーチテクノロジー社 (TTS商品を扱っていた) を買収。1997年には、STT分野ではKurzweil Applied Intelligenceを、MT分野では、Trantex社 (フィンランド)、Wordwork社 (スウェーデン)、Kermit社 (イタリア) GMS社 (ドイツ) をそれぞれ買収した。1998年に入っても、TTS分野のTikSoft社など4社、またMT分野では、Heitmann社を買収している。L&H社の狙いは、従来から強みのあったSTT技術を多言語化して、さらにMT技術に強い企業を買収により、STTとMTとTTSの多言語化を果たして、クライアントのウェブビジネスをサポートすることである²⁸。

5. 日本参入の可能性

food.comの躍進をみると、米国以外の市場への展開も考えられるが、日本市場には受け入れられるだろうか。ケータリングあるいはテイクアウトを利用する日本人の割合は、米国人と比べればかなり少ないであろう。ただし、それも居住地域、家族構成や就労状況 (共働き) によって違うであろう。例えば、都市部とそれ以外では生活パターンも異なる。少なくとも筆者の周りには日常的にデリバリーやテイクアウトを利用しているものはいない。しかし、昨今の宅配ピザ業者の急成長ぶりをみると、日本にもそろそろデリバリーで頻りに食事をする時代が目前にきていると感ぜられる。日本では都市部を中心にfood.comのマーケットが広がるのではないか。

²⁶ <http://rivendel.com/ric/resources/mt.html>

²⁷ telephony application programming interface, telephony server application programming interface とともにAT&TとNovell が中心となって取りまとめた。

²⁸ <http://www.lhs.com/company/lhar98.pdf>

米国のレストランなどの飲食店では、パソコンなどの情報機器に対するリテラシーは高くない。ニューヨークのマンハッタンのレストランでは、大抵黒人やヒスパニックあるいは中国系など白人以外の人種が給仕をしていることが多い。また、統計的に、非白人は教育程度や所得も白人にくらべて相対的に低い。したがって、food.comのようなビジネスはレストラン側のニーズにきわめてよく合致していると言えよう。給仕を勤められるのであるから英語に対するリテラシーは問題なくとも、PCリテラシーの低いスタッフの多いレストランにPCあるいはITの恩恵をもたらすには、旧来のメディアつまり電話を組み込んだ仕組みがよいことは想像に難くない。

日本でのリテラシーはどうであろうか。まず、一般にPCリテラシーは若い世代ほど高いと言えよう。しかし、高年齢になればなるほど、PCなどの情報機器を苦手とする人が多いであろう。従って、レストランにとっては、food.comのサービスを受け入れる余地はあろう。また、food.comの最大の導入メリットである顧客に対する24時間サービスを低価格で行えるようになるメリットもあろう。

6. food.comの課題

まず、営業費の増大が挙げられよう。food.comのサイトで自宅近辺のレストランがそもそもなかったり、仮にあったとしても気に入ったレストランが見つからなかった場合、その顧客を取り戻すのは難しい。そこで、food.comは利用できるレストランの拡大に努めている。現時点では、レストラン数拡大のペースが顧客流出のペースを抑えているようだが、今後逆の悪循環を起こすことがないように、food.comのレストラン拡大にかけなければならない営業費は増え続けるだろう。

顧客とレストランへの認知度を高めるための広告費の高騰も懸念される。food.comによれば、前述のロスアンゼルスタイムズの他に、シカゴトリビューンやワシントンポストにも、メディアパートナーとしてfood.comのバナーを出しているとしている²⁹。しかし、筆者が時間帯をかえて両メディアのサイトをチェックしても、food.comのバナーを見つけられなかった。また、この他の比較的大手のメディアを新たにパートナーにしたという報道も見られない。food.comの財務データは公表されていないので断言できないが、戦略的に取捨選択してロスアンゼルスタイムズやAOLをメディアパートナーとしたというよりは、バナー広告費の高騰により現在のようにならざるを得なかったと言えよう³⁰。

food.comのサービスに対する顧客満足度が、レストランが顧客に提供するサービスのレベルにほとんど依存していることは、危険を内包しているといえる。例えば、food.comからの注文が少ないレストランの場合などは、不慣れなために注文を間違ったり、あるいは顧客に対する配達員の対応が悪いことも考えられる。いかにfood.comのウェブサイトがきれいで、決済には最新の暗号化技術を用いようとも、最後には「ひと」が介在する。現在のように、レストラン数拡大をひたすら続けていくのであれば、この危険は増大するであろう。

レストラン側が顧客の生の要望なり現状を把握できなくなる可能性もある。

²⁹ <http://www.food.com/cgi-bin/WebObjects/>

³⁰ 全く別の分析として、その将来性を見込んで世界最大のISPであるAOLがfood.comを買収することも考えられる。その布石として現在のようにAOL依存が深まっているのかもしれない。

food.comは顧客からの注文を電子メールで処理することにより、業務の効率化を行う反面、顧客からのレストランに対する「余計な」情報を排除してしまうことになる。Aという料理を本当は食べたいのだが、AしかないのでAを仕方なく注文した顧客の声はレストランには届かない。電話で問い合わせる時や注文を受ける時に、顧客からの情報つまりマーケット・インフォメーションを得にくくなるだろう³¹。

7. 結び

将来の広帯域光ファイバー時代には、すべての端末が今より多くのインテリジェンスをもつようになるであろう。これは現在のような通信キャリアによる中央指令型のネットワークから、無数のPC類似の機器を端末とする自律分散型のネットワークへの移行によるものである。food.comは事情は何であれ、将来型のネットワーク（サービス）への過渡期に、現れるべくして現れたとも言えよう。food.comは先進的なユーザーを対象としてPCから受けられるサービスを提供する面と、その一方で、まだ情報化の進んでいないレストランに対しては、旧来型の電話とファックスで対応する可能性を提供した。言い換えれば、電子データをアナログデータに劣化させているので、つまり、food.comのビジネスは、情報化の新旧両側面を具備しており、まさに過渡期のサービスと位置付けられる³²。

TTS、STTとMT技術は急速に融合している。また、商品化も進んでおり、ウェブビジネスを行う企業も投資に見合うリターンを十分に期待できるようになっている。それは、顧客対応などのオペレーションに必要な人的稼働を削減することに始まり、food.comのような新たなビジネスを生むであろう。融合する技術によって多言語化された商品により、インターネット本来の特質を生かした世界規模の新たなビジネスが現れるのは確実である³³。

鈴木 淳弘（すずきあつひろ）

国際大学グローバル・コミュニケーション・センター主任研究員

³¹ ひとつの解決策がオンライン書籍販売のアマゾンのように、顧客からのフィードバックをとり入れる仕組みを導入することであろう。あるレストランを使った顧客の評価をフィードバックできれば、他の顧客にとって店選びの基礎になると同時に、レストラン側にとっても顧客動向を把握できるようになる。

³² アナログの音声電話をメインに考えてきた通信キャリアにとっても、データ（電子メール）を音声トラフィックに回帰させるメリットがある。

³³ 本稿に対して、GLOCOMの公文所長をはじめ、青柳、竹田、上村、土屋各先生には貴重なアドバイスをいただいた。深く感謝の意を表したい。