

音声合成（AI アナウンサー）と放送の現在

柴田 秀一*

はじめに

「音声合成」とは

音声合成の仕方

エフエム ワカヤマの例

TBS ラジオの例

NHK の例

南海放送ラジオの例

まとめ

はじめに

本稿を執筆中に中国・新華社通信が AI キャスターのお披露目をしたとのニュースが入ってきた。(2018/11/07)⁽¹⁾ 実際に新華社に所属する男性キャスターがモデルになり、CG 合成風の画面動画では、喋るとそれに合わせて唇が開閉してリップ・シンクもする。英語版、中国語版があるが、小職は、英語も得意ではなく、中国語も全くできないが、この AI アナウンサーは「緊張気味に話した。」との記事もある位なので、とても本物との区別がつかない、というほどではないものの、何を言っているかわからないということはないようだ。

2017 (H29) 年、日本の放送局では、人工的に音声を発する機械（コンピューター）を使った放送の実験や、実用化が相次いでなされた。いわば、人工音声装置での放送元年とも言える動きが出た年になった。

愛媛県の南海放送ラジオでは、1 か月に亘る実験放送、和歌山県 FM 和歌山のレギュラー放送ニュース、東京 TBS ラジオの月曜～木曜の番組「好奇心家族」。静岡放送テレビでの夕方ニュースワイドでの実証実験、更に翌年 NHK テレビ夜のニュースでは、週 1 回レギュラーで音声合成アナウンサーがニュースを読んだ。

そこで、本文では、人工的な音声を使った放送が、日本のテレビ、ラジオでどのように放送されているかを探った。小職は、TBS テレビという放送局で 36 年アナウンサーとしてニュースを中心に係り、現在も週に何度かテレビ、ラジオ放送に携わっている。その経験から、

- ① 音声合成の完成度と放送での利用
- ② 働き方改革としての利便性

に注目して放送目的での音声合成の現状と今後、進む方向、実際の放送組織、放送の仕方との関

*しばた しゅういち 日本大学法学部新聞学科 教授

係、特に「AI〇〇」と言われて、人間にとって代わられ、「仕事の間を奪われる」との懸念が絶えないことについても、労働の仕方、働き方改革と共に考察する。

本文では、実際に音声合成による放送をした NHK（テレビ）、エフエム ワカヤマ、南海放送（ラジオ実証実験）、TBS ラジオ担当者に、実際の放送のやり方について話を聞きまとめた。

尚、可能な限り音声合成の「声」を HP やクラウドなどを使い「注」でアクセスできるようにしたので、是非聞いてみてほしい。「まだまだだ」とお思いになるか、「ここまでは来ているのか」とお思いになるか。お聞きになった方は、どのように判断されるだろうか。

「音声合成」とは

スマートフォンで、「シリ！（Speech Interpretation and Recognition Interface, の略 Siri）」（Apple・「iPhone」）と呼びかけると反応する人工音声応答システムが既にお馴染みで、スマートスピーカーでは「オクケー、グーグル」（Google・Android 端末向け）で様々な検索をして答えてくれる。家庭においても「アレクサ！」（Amazon.com「Alexa」）と呼びかけ、電気やテレビが灯いたり、ニュースを読み上げたり（このアレクサで、後に説明するNHKは、合成音声「ヨミ子」がニュースを読んでいる）。車に乗っても、「メルセデス、車内温度下げて」（ベンツ応答システム）でお望み通りの温度になり、人工音声はきちんと返事をし、人間のように気分が無視はしない。

これらは「ヒト」の声を聴いて、単語として認識し、質問に対する最も「確からしい解答」を「音声化」して発信するシステムで、この「答えの音声化」の部分で『人工音声、音声合成、合成音声、AI アナウンサー、バーチャル・アナウンサー、アバターアナウンサー』等と様々な名前を付けて表している。

名前は色々ついているが、「ヒト」の声ではなく人工的に作った音で、「ヒト」の声に近い音を出すもので、研究者は、「音声認識」と「音声発信」を合わせて「音声合成」（speech synthesis、voice synthesis、text to speech）と言っている。

「音声合成」開発を続けている研究者、倉田宜典氏（ソニー株式会社 ブランドデザインプラットフォーム UX・事業開発部門 シニアテクニカルプロデューサー、氏は、ロボット犬アイボの開発にも携わっている）に聞いた、「音声合成」を理解するための基礎知識とは以下のようなものである。

音声合成の仕方

現在放送で使われている「音声合成」システムは、「原稿を音声化する」というものがある。

それは、例えば、楽器の自動演奏装置（ピアノ等）に楽譜を読み込ませて、その通り演奏させるものと似ている。楽譜が原稿で、出る音が楽器の音ではなく「ヒト」の声に近い音であることが違う点だ。音は「シンセサイザー（synthesizer）」でいかようにも変化させることができると考えると、「ヒト」の音声に近い音が出る事が理解しやすい。

演奏させるための指示記号である「音符」が原稿の言葉と音の間や、日本語アクセントの高低を表す記号と共に指示されて（「中間言語」といういわば指示書）、コンピューターに命令として読み込まれ、それがアプリケーションに従って「ヒト」の声に近い音を出していくわけである。

例えば天気予報なら、「音声合成」に喋らせたい制作者は、天気予報文を PC で打ち込み「中間言語」を作る。

「中間言語」とは、テキストを「音声合成」が読み上げる際に、日本語の高低アクセントをつけたり、間を開けたりするための情報を記号化したものであり、いわば自動演奏の指示書、楽譜だ。これはアプリケーションを動かすための指示で、その指示に従って（音符の音の長さや高低に従って自動演奏する様に）「ヒト」の「声」に似た音を出すのである。

これは音のみだが、テレビの場合はキャラクターの動作やリップシンク、カメラワーク等が選べ、それに、ビデオを加えて同時に流せるようにする。

こうした「中間言語」で出来た音を流してみても、アクセントや間の取り方が違えば、「中間言語」を修正できることが特徴で、放送するまでに、より読みを自然にすることができる。

この「音声合成」が構築され、今日に至るには、スマートフォンの発達が大きいそうだ。スマートフォンでの会話、検索命令等の会話資料は匿名情報でメタデータとして資料利用された「shiri」や「OK Google」また、スマートスピーカーも同様である。そうしたヒトの会話資料が、より、自然にしゃべる「音声合成」を構築するために役立ったという。

倉田氏は、株式会社 共同通信デジタル社が提供している「音声合成・アバターエージェントサービス・バーチャルアナウンサー 沢村 碧」の開発に携わっている。

「沢村 碧」はテレビ、ラジオの放送局用に提供している「音声合成」バーチャルアナウンサーとして、ニュースや天気の情報原稿を PC 用アプリケーションに入れると、「沢村 碧」のキャラクター（設定は 25 歳、7 月 24 日茨城県つくば市生まれ、身長 158cm）の音（声）で、一定の速度で読み上げるもので、番組制作者が、映像機器や PC、CG 画像に詳しい知識がなくても、キャラクター「沢村 碧」がニュースを読みあげる動画ファイルを簡単に作ることができ、リップシンクや表情も付きながら放送できるというシステムである。



株式会社共同通信デジタル ソニー株式会社

この「沢村 碧」が出演して、愛媛の南海放送ラジオ、静岡の SBS 静岡放送テレビなどで、2017 年に実験放送が行われ、同年 10 月から翌年 3 月まで TBS ラジオでも月曜から木曜まで夜の 4 時間ワイドにも出演した。

沢村碧の声⁽²⁾ (URL でご覧、お聞きください。)

エフエム ワカヤマの例

エフエム ワカヤマでは、2017年7月から、人工知能アナウンサー「ナナコ」の放送を開始した。早朝、深夜を中心に、最新のニュースと天気予報を伝えている。「ナナコ」は、アマゾンウェブサービス（AWS）の人工知能「Amazon Polly」を使った「音声合成」によるアナウンサーであり「ヒト」であれば女声である。男声には「八太郎（はちたろう）」と名前がついている。

エフエム ワカヤマでは人工知能アナウンサー「ナナコ」で以下の放送をしている。

毎日 [7:08] AI ナナコニュース [7:18] AI ナナコ天気予報 [16:01] AI ナナコニュース
[21:01] AI ナナコニュース [21:19] AI ナナコ天気予報（月～木）
[23:54] AI ナナコ天気予報（番組表に載る主なもの⁽³⁾）

エフエム ワカヤマ⁽⁴⁾は特定非営利活動法人で、和歌山県和歌山市に本拠を持つFM放送の特定地上基幹放送事業者である。いわゆるコミュニティ放送⁽⁵⁾で、全国322局（2018/12現在・日本コミュニティ放送協会）のうち、和歌山県には協会加盟4局ある放送局の1つ。周波数87.7MHzの語呂合わせで「バナナFM」と愛称をつけている。

2008年4月1日放送開始、2012年には和歌山市と防災協定を結び、防災緊急放送が市から放送された場合、エフエム和歌山の放送が強制的に切り替わり防災無線と同じ内容が放送される。防災FM放送局となる。出力20KW、聴取可能地域是和歌山市全域、岩出市の一部、紀の川市の一部、海南市の一部で、人口集中エリアに向けた放送として45万人に届けられるとしている。



エフエム ワカヤマ 外観



第1スタジオ

エフエムワカヤマで「音声合成 ナナコ」を使用し放送することになったのは、上記の地域災害放送が目的である。特に、大規模災害時に役立つラジオを目指している。東日本大震災後の臨時災害放送局の実際の放送形態や放送内容を踏まえて、災害が起きてから電波が割り振られて放送が始まるという法律上の欠点を補い、災害のない普段から、放送を聴いて周波数を覚えてもらい、いざ災害というときに直ぐに聞いてもらえる放送局にしたいとの担当者、エフエムワカヤマ クロスメディア局長の山口誠二氏の強い思いがあった。

臨時災害放送局（臨災局）では、被災後に周波数が決められる。災害での混乱が続く中、それを被災者に告知して、知ってもらった上で周波数を合わせない限り放送を聴けない。また、FMラジ

オを所有していなければ聞けないし、普段から車などで聞いていないと、馴染みがなく聞こうという気にならない。「臨災局」は放送を経験していない人が、プロから教わったり、見よう見まねで放送している。局を立ち上げ運営すること自体が大変なことであるが、そうした欠点を補い、普段から聞いている局で災害の折情報をとれることがエフエム和歌山の考え方である。

更に、災害放送で威力を発揮するのは「音声合成」による放送だという。もし大規模災害が起これば、コミュニティFMは、地域住民のために、長期間にわたって災害に関する情報や地域の生活情報を発信し続けることになる。だが、深夜・早朝で放送を担当するアナウンサーすら確保できない可能性もある。そんなときに「自動音声」による放送ができれば、たとえ原稿を読める人間がいなくても、情報を入れさえすれば24時間放送が可能である。

「ナナコ」ニュースや天気予報は、スタジオを使用しないでコンピューターから出している。

自らがプログラマーだった山口氏はエフエムワカヤマで2つのシステムを作った。

一つは設定した時間通りにニュースや天気を流すシステム（オンタイム・プレーヤー 2017/7/1～）で、実際に「ナナコ」で放送をしているシステムである。

例えば、ニュースの原稿は契約している読売新聞社からWebで原稿が送られてくるが、受信したときにプログラムが文書を解析してデータベースに入れていき、時間になったらそのデータベースから選んで出していく。

ニュースを選択するプログラムを作っていて、一番新しい5件を再生する。新聞社が付けている星3、2、1という重要度の高いもの順で、株と為替があった場合には最後の項目にする。全国ニュースは出す必要はない。また殺人・殺害は自動的に削除、しかし、和歌山の事件事故はやる。選手の個人名が多く、読み間違える可能性のあるスポーツニュースは扱わない。局としては、悲しいニュースはできるだけ止めて、科学技術の進歩など明るいニュースを出すという編集方針があるので、これらを自動的に選別して放送することが可能である。

また、天気は、予報に使う語彙はニュースと比べるとはるかに少なく、晴れ、雨、降水確率、注意報警報等言い回しが決まっているものが多いので、文章を定型化しやすいのだ。

もう一つは災害時に最新情報が配信されるとそれを合成音声化して放送し、終わると音楽が一曲流れるもの。情報が更新されるとまた、同じように放送され曲もかかるので、繰り返し放送できることから、音楽用語の「ダカーポ Da Capo (2017/9/1～)」(はじめに戻る意)と言うシステムである。

このシステムは、情報（文字情報として）が入りさえすれば、たとえ局内が無人であってもシステムは動くのである。情報は、コンピューター通信で繋がっている新聞社や気象情報の会社、役所から入る文字情報を「ナナコ」が音声化できるように自動的にテキストに作るシステムがあるので更新できる。このあたりはプログラマーだった山口氏の力量によるところが大きい。

しかし、災害に関する情報は、どれが確実な情報か絶えず確かめる必要があり、誤報を流すことは命に拘わるので元々自動とせず、警報、注意報、避難等の情報は一つ一つ確認して出していると言う。

山口氏はこうしたシステムを他の放送局にも利用できるようにしたいと考えている。また、問い合わせも来ている。

特に人手不足の現場に使えるのではないかという。アナウンサー等、喋り手の人員が削減される事を恐れる現場とは、逆の発想である。

アマゾン ポリー (Amazon Polly) 採用の経緯

エフエム ワカヤマの「人工知能アナウンサー ナナコ」は「アマゾン ポリー」(Amazon Polly)⁽⁶⁾のシステムを採用している。

もともと災害放送を自動的にできないかを考えていた山口氏は、3年前の2015年 HOYA (音声合成会社) でレベルの高い音声合成を達成していたのを知った。しかし、当時、初期費用で80万円くらいし、月々の費用に月額7~8万円だったため費用が掛かりすぎて断念した。

2016年にAWS (Amazon Web Service) の人工知能を使ったPOLLYが(テキストを日本語で使用しても)100万文字で4ドルだった。1年使っても100万は文字いかないので、費用が大変安かった。

最初に検討したHOYAのシステムはAPI (application programming interface) というシステムを使わなくては行けないが、独自の関数などを使っていて、他にどこも使っていないので比べられなくわからない。プログラマーが使える人工知能(AI)を使った音声合成では「AWS」、「IBMワトソン」、「Microsoft アジュール」、の3つで、当時グーグルはサービス化していなかったもので、聞き比べたところPollyの音(声)が一番良かったという。3つとも言葉(漢字の読み)の間違ひはあるが、「機械学習」すればよいと考えた山口氏は、はラジオ放送に使って大丈夫と思えるのはAmazonと判断した。

エフエム ワカヤマが、何故「AIアナウンサー」と呼んでいるか。

いままでの「音声合成」は一回基本ソフトをインストールするとそのまま、使い手が同じ単語や文章のアクセント、イントネーションが間違っているのも、変えていきたい場合、それを指示し、手直ししないと行けないが、「Amazon Polly」は、それを自分で学んで修正していくので、「AI」なのだという。そのため2017年の放送開始から1年以上経っているが初期と今では明らかに「読み」は進歩しているのだが、どうやって「機械学習」しているのかは、「Amazon」は秘密として明らかにしないそうだ。これが、機械学習が「ブラックボックス」と言われる所以である。

つまり、「ヒト」の声に近い音で、「ヒト」が読むアクセントとイントネーションに、機械が自ら学習して近づけていくので「AIアナウンサー」なのである。

エフエム ワカヤマでは毎日ニュース・天気をナナコが読むことによって、「ナナコ」という「音声合成」キャラクターは定着しつつある。「聞き苦しいから止めろ」といった意見はリスナーからはほとんど来ないそうで、逆に、意外と「わかる」、「聞ける」と受け止められている。

災害放送での使用

では災害放送として、緊急情報に対する運用実績は、どうか。

2017年秋に3回台風が来たときは、この仕組みによってディレクター1人によって放送対応で来たという。2017年台風18号の影響で、和歌山市内では最大4,400世帯が停電する被害が発生。台風通過中の約5時間連続で天気、災害情報を「音声合成」「ナナコ」で放送。提供される情報は随時更新されて放送された。情報はどう取っているか。避難勧告は、対象地域、その避難場所、今

何人がそこに避難しているのかまで調べた。気象庁のウェブサイトで気象情報を確認、電車の遅延や渋滞等交通情報はFAXで届き、停電は関西電力の停電地区と軒数を確認、原稿にまとめる。アップデートされた情報は原稿を更新する。電話問い合わせもしないと正しい情報は得られない。台風の情報収集は毎年のもので慣れているので問題はなかったというのは、潮岬がある台風銀座の異名があった和歌山特有の経験が生きている。

例として、2018年にも影響があった台風21号（2018年9月4日）と西日本豪雨（2018年7月5日）の際の放送をクラウドにアップしたのでお聞き頂きたい。⁽⁷⁾

内容は以下の通り。

『(西日本豪雨による被害 2018年7月5日5時)

和歌山市内の道路の情報です。田中町の地下道は通行止めになっています。

手平の地下道は通行止めになっています。カセイ橋は通行止めになっています。

道路冠水の情報です。内原・パチンコ銀河からファミリーマートの間、松江東・衛生研究所付近 田尻 塩見橋北側 府中盲学校北側 紀三井寺看護学校付近 紀三井寺団地南側道路 このほかにも冠水や通行止めがある可能性があります。通行には十分注意し不要な外出は控えてください。

(台風21号による停電被害 2018年9月4日13時)

関西電力によりますと、現在、和歌山市で約3690軒の停電が発生しています。朝日で約10軒、内原で約670軒、北中島1丁目で約420軒、小雑賀3丁目で約130軒、手平1丁目で約10軒、手平3丁目で約110軒、手平4丁目で約40軒、手平5丁目で約230軒、西浜で約1280軒、広原で約30軒、冬野で約770軒の停電が発生しています。』

アンダーラインを付したところはアクセントが違うと思われるところ。地名のアクセントはその地方独特のところがあるので、必ずしも間違っているとは言えないが、市役所にも確認したところ停電情報の「広原」を「ヒロワラ」と読む間違いがあった。しかし、そういうところを除いては市民に情報を出すという点で見れば、全国にではない地域の災害放送としては、流さないよりはるかに良いと思われる。

TBS ラジオの例

TBS ラジオは東京港区にあるAMラジオ放送局でありTBSテレビとともにホールディングスグループ会社の一つであり、民間放送ラジオネットワークJRNのキーステーションである。

TBS ラジオは「THE FROGMAN SHOW A.I. 共存ラジオ 好奇心家族」を2017年10月3日(火)からスタートさせた(2018年3月31日終了)。フラッシュアニメ「秘密結社 鷹の爪」の生みの親で、それに登場するすべてのキャラクターの声を演じ分けるマルチクリエイター「FROGMAN (フログマン)」が、“A.I. (人工知能)”とラジオで史上初の掛け合いをするというもので週4日毎週火曜日～金曜日 17:50～22:00 (4時間10分)というワイド番組である(関東以外の放送局とも一部ネットワークで流れていた)。この中には、株式会社共同通信デジタルが提供するアバターエージェント「沢村 碧」と、日本語を話すペット=人工知能犬「ドッチ君」(株式会社エーアイ+Jetrun テクノロジー株式会社)⁽⁸⁾⁽⁹⁾が出演している。

番組内では AI と、パーソナリティーやアシスタントの女性 TBS アナウンサーとが会話をしたり、天気予報を伝えたり、質問や検索のやり取りをし、AI がリスナーからのツイッターやメールを読んだり、今日の気分の曲を自動選曲したりした。



↑ TBS ラジオ好奇心家族

何故ラジオ番組に AI を登場させようと思ったのか、この番組担当プロデューサー・ラジオ編成局 橋本吉史氏に聞いた。

「AI 時代にふさわしい番組にしたかった。番組開始前スマートスピーカーがブームになりはじめた。世の中は音声入力（機械に語りかける）になっている。同時にラジオ的コミュニケーションが主流になると考えた。ラジオが何もしないのはおかしい。半年限定の番組だったが、番組は、AI を知る、AI を体感する番組、生活の中で AI を使うとこんな感じと分かってもらいたい。

しかし、番組中、AI ペットの「ドッチ君」に話しかける演出をするが、すぐ答えが出ず、意外にうまくいかない。だから、AI 共存は「まだ、今このくらい不便だよ」でいいと考えた。

バーチャルアナウンサー「沢村 碧」に天気を読ませているが、「アナウンサーを凌駕する」などとは思っていない。それは無理である。例えば、アナウンサーが休んだなどに、決まり文言を読ませるのを「共存するラジオ」として、「職場を奪う」のではなく「働き方改革を助ける」存在であると考えている。」と語っている。

この場合もエフエム ワカヤマと同様 AI ペット犬に「ドッチ君」と名前を付け、キャラクター化を図っている。呼びかけて答えを聞く「スマートスピーカー」と同様の機能があるが、「ドッチ君」と名前を付けることによって、もしうまく答えられなくても「がんばれ!」「次はできる」など、リスナーが擬人化し、感情を共有するという効果も現れる。

更に、株式会社ソケット⁽¹⁰⁾の提供する AI 自動選曲機によって、「赤の曲」といえば、「赤」の歌詞が含まれていたり、情熱的リズムの曲、などを選曲して演奏する機能も使われて、リスナーのリクエストに応えた。

では、最先端をいって、スマートスピーカーを使いこなしている人がリスナーなのかというと、大半のリスナーはファミリー層や子供、高齢者で、概して保守的で、スマートスピーカーは「持っていない」、「高い」とか、「そんなに言うんなら、買ってみようかなあ」の程度だという橋本氏は「それでいいのではないか。まだ、そんなもので、今、生活の中で使ったらどうなるか体験しても

らうことでよく、紹介使われるようになれば、自然に生活に入ってくることになると思う」と話す。

NHK の例

ニュースとスポーツ ロボット実況



NHK ニュース ヨミ子

AI アナウンサー ヨミ子

NHK では、テレビニュース番組に 2018 年 4 月から「AI アナウンサーヨミ子」を出演させている。

ニュースヨミ子のホームページ「ヨミ子の部屋」⁽¹¹⁾によるヨミ子の紹介をまとめると、ヨミ子は「ニュースのリポーターとして採用され、NHK「ニュースチェック 11」（夜 11 時 10 分～総合テレビ月～金）の毎週水曜「ヨミ子のニュース」に出演。それ以外、テレビに出ていない時はスマホの中に住んでいて AR（Augmented Reality・拡張現実＝現実情報にデジタル合成等で作られた情報を加え人間の現実認識を強化する）でいつでもどこでもスマホに登場し、風景や人物と一緒に写真を撮ることもでき、HP 上で聞きたいニュースの項目を選ぶと、「ヨミ子」が最新のものを読み上げる。また、Google Home、Amazon Echo など対応するスマートスピーカーに向かって Google アシスタント「ヨミ子につなげて！ ヨミ子と話す！」、Amazon アレクサ「ヨミ子をスタート！ ヨミ子を呼んで！」などというようにヨミ子が登場する。

この HP（前述の「ヨミ子」の紹介が書かれている「ヨミ子の部屋」）では「ヨミ子」が読むニュース（15 秒～30 秒程度のもの 6～7 本）が更新され、最新のものでクリックすると聞けるようになっている。また、「ニュースチェック 11」で出てくるアニメ動画も再生することができる。ヨミ子についても、「音声合成の仕方」の項で書いた通り、ニュース原稿から「中間言語」作り、それを一度音に出して見て、調整を加える方式だ。音（声）のベースになっているのも 1 人の女性アナウンサーで、何十時間か様々なニュースを読んで、ヨミ子の音のベースの「読み方や間」の為にニュース学習資料を作り、半年間ほど機械学習を繰り返したという。NHK の場合は機械学習を間に挟んでいるので、エフエム ワカヤマの Amazon Polly に近い方式だ。

再生してみると、原稿は記者が書いているので、内容は問題ないもののショート・ニュースの

為、一文が長く「。」が少ないので、現役アナウンサーが読んでもやや難しい。そこで、「ヨミ子」のアクセントやイントネーションにやや難が残る。

(この項の「ヨミ子」のニュース音声)⁽¹²⁾

例えば、「(大阪寝屋川市の二人子供殺害事件 2018年12月19日43秒)：三年前大阪寝屋川市で中学1年の少年と少女を殺害した罪に問われ、死刑を求刑されている48歳の男への判決の言い渡しが、大阪地方裁判所で行われています。裁判所はこれまでに、被告が二人を殺害したと認定しました。結論に当たる主文は最後に言い渡されます。」

このニュースの場合、まず、アンダーラインが引いてある

「求刑されている」の「いる」が「ヨミ子」は「いゝる」と頭高(「い」が高い)になっている。普通は平板(二音目以降が高く、助詞が付く場合助詞まで高い)だが、「48歳の男に」掛かる「句」と考えて「いゝる」と同じ高さで読む。

「言い渡しが」は「言い渡し」のアクセントが平板なので次の格助詞「が」は音が低くならず、「いいわたしゝが」と同じ高さの音になる。

「行われています」では、「行／わゝれてい／まゝす」とヨミ子は発音しているが、「お／こなわれてい／まゝす」となる。

「被告が」の「被告」は平板なので、続く助詞「が」は下がらず「ひ／こくゝが」

「認定しました。」は、「ヨミ子」は「にんていしゝました」と「し」の後で音が低くなっているが、「に／んていし／まゝした」と平板の「認定」と中高(単語の2音目以後いくつかの音が高くなり最終音までには下がる)「しました」が付いた形になる。

(尚、／は記号の次の音が上がる、ゝは記号の次の音が下がる、ゝは、音はそのままの高さの意)

更に、「ヨミ子の窓口」としてニュースの発音やARの動作など、何かおかしいなと気づいた人が連絡できるようになっているのも、ヨミ子の読む力を上げるためのものか。

更に、「ヨミ子」がニュースに出演する前、2018年2月に開催された平昌オリンピックで、世界初となる音声合成による自動実況「ロボット実況」を放送し、17競技で毎日1本ずつ32時間放送した。この開発技術がヨミ子の同年4月からのニュースへの出演を可能にした。

ロボット実況

平昌オリンピックでは「音声合成」の「ロボット実況」で、NHKオンライン特設サイトおよびNHKハイブリッドキャストで、ソリ系の3種目(ボブスレー、リュージュ、スケルトン)とアイスホッケー、カーリングを対象に、試合の映像に自動で実況と字幕を付けて放送された。

これまでは、インターネットで、実況の無い生中継動画を配信していたが、今回、一部の競技で実況付きの動画配信をした。

ロボット実況とは、人間のアナウンサーに代り、コンピューターが自動でスポーツ競技を実況するサービスだ。視覚障害者をはじめ、より多くの人にスポーツ中継を十分に楽しんでもらう目的で作られた。大会の主催者が配信する競技データから、即時に実況内容を作成し、「合成音声」が読み上げる。同時に字幕もつけている。

NHK 技術研究所スマートプロダクション研究部主任研究員 熊野 正氏によると、「ヨミ子」「ロボット実況」はもともと視覚障害者の支援放送の研究の為にできたものであった。目の不自由な視聴者がスポーツ競技をテレビで楽しむためのものである。ドラマでやっている解説放送（登場人物の動きや感情などが、分かりやすいように説明し、セリフの合間に音声化したもの）

リオ・オリンピックで、そうした実況の合間に状況解説的に出して試したものを、平昌冬季オリンピックでは、実況アナウンサーの喋りは無く、「音声合成」と現場の音と映像のみで構成した。

小職が見たのはデモテープ（放送したものは別のもので、イベント用で見せるために作ったがもので、オリンピック映像が含まれるので公開されていない）だった。スケルトン（うつ伏せで滑る金属製の1人乗りソリ）は、実際のアナウンサーの放送とそれほど大きく変わらないのではないか、という出来だった。

一部言葉に起こしてみると、

『スタートしました。第1チェックポイント4秒62で通過 現在時速48.6km。第2チェックポイント13秒37で通過 現在時速101.8km。第3チェックポイント22秒74で通過。1秒44上回って現在トップ 第4チェックポイント32秒89で通過。第5チェックポイント39秒26で通過。現在時速125.2km 韓国の〇〇選手50秒02でゴールしました。現在トップです。』

民間放送のスポーツ中継を聞きなれた人間からするととても沈着冷静で、数字のアクセントはほとんど狂わないが、日本人の選手でも他国の選手でも、世界記録が出るような状況でもこの感じだと冷静すぎるかと感じる。それでも実況用に機械学習を改めてしたそうである。視聴者からは、冷静に聞けて良かったという反応があったという。

こうした実況はどうして出来るのかというと、まず、放送する前に出場者等の名簿と滑走順をあらかじめ打ち込み、「スタートしました」「第1チェックポイントです」などといったテンプレート（雛形）を作る。第1～第5チェックポイントに来たら中間タイムを音声化する形のテンプレートも用意する。するとスタートし、情報が入ったら、テンプレートが動き、上記のような音声化ができる。

これは、オリンピック放送機構（OBS：Olympic Broadcasting Services）がODF（Olympic Data Feed）と呼ばれる競技のリアルタイムデータを各放送局に配信しているからでありこれを有効利用して音声化に生かせるのだ。特にスケルトンは競技時間が短いので、こうした実況には適しているかもしれない。1人で完結する競技はロボット実況しやすいとみられる。

このほかに、対戦競技である、アイスホッケーの一部は次の通り『スウェーデン対日本、現在第1ピリオド両チーム無得点。・・・（中略）・・・スウェーデン〇〇選手のペナルティー インターフェアランスです。パックを持っていない選手を妨害する反則です。2分間の退場です。スウェーデン4人対日本5人になります。』と実況しているがこちらはデータの到着時間によっては実況に少し遅れが生じる。

また、種目によっては、事前に告知されたタイムテーブル（選手紹介などをテンプレート化していても）通りに行われる種目と全くそうならない種目もある（アイスホッケーもそうらしい）。そうすると画像と、音がずれて実況にならない場合もあるので、今のところスポーツ実況は種目を選んでということだと思う。ただ2020年の東京オリンピックでは、ロボット実況はより規模が大きくなると予想されている。

更に、顔認証ができれば、例えばアイスホッケーでゴールをした選手を言うのは、可能なのではないかと聞いたが、「画像解析」について、今は対応していないとの事で、後の研究でどうなるか。理論的には可能だという。

南海放送ラジオの例

南海放送株式会社は、テレビ・AMラジオ兼営放送局で愛媛県松山市に本社があり、テレビは日本テレビ系列、AMラジオは JRN、NRN⁽¹³⁾ の両方の系列に属している。

南海放送では2017年に「音声合成」による放送の実証実験をラジオ放送で行った。

プレスリリース⁽¹⁴⁾によると、

『2017年7月1日から1か月間、株式会社共同通信デジタルと共同で、バーチャルアナウンサーがニュース原稿などを自動音声で読み上げる実証実験を行う。通常、ニュースはアナウンサーが読むが、今回の実験では音声合成技術を利用し、ニュース原稿を文字情報から音声に自動変換してバーチャルアナウンサーが伝える。実験では文字情報の入力作業にかかる負担や、ラジオリスナーがニュース内容を十分に理解できるかどうかなど実用化の可能性を探る。』

実証実験を行う番組と放送時間は、「正岡省吾の RADIO MONSTER!」（毎週土曜日 午前8時30分～11時10分）内で随時放送（初回は2017年7月1日）』



RADIO MONSTER
プレゼンター 正岡省吾氏



バーチャルアナウンサー
沢村碧

南海放送は、株式会社共同通信デジタルとソニーのシステムである「沢村 碧」（本文の「音声合成の仕方」の項で述べたアバターエージェントサービス）をラジオワイド番組に出演させる方法で、7月1日から1か月「1日、8日、15日、22日、29日」午前10時半ごろから計5回放送した。内容は、「沢村 碧の天気予報」で人工音声沢村碧が天気予報を読むというもので、番組パーソナリティーの正村省吾アナウンサーが呼びかけ、それに「沢村 碧」が答えるという形式で、音声合成だけで天気を放送するのではないラジオのワイド番組ならではの工夫をした。

担当した、南海放送メディア本部ラジオ局ラジオ業務部副部長の三瀬雄一氏に話を聞いた。

「アバターエージェントサービスを利用したラジオ放送は全国で初の試みだった。」

このシステムがあることは2年ほど前に知った。やり方は、原稿化したものを音声に出すもので、「中間言語」というカタカナと記号ばかりの原稿にする。例えば、記号を入れることによってアクセントで音を上げる場合ピリオドを2つ付けたら半音上がる、もっと上げたいときは%入れるといった指示で出来る。放送してみるとリスナー反響は好意的なものが多かった。「初々しい」など

もあり、一日10通～20通くらいはEメール、便りが反響としてきた。初回（7月1日）ラジコ⁽¹⁵⁾のユニークリスナー数⁽¹⁶⁾（ラジコを聞いているユニークユーザー数）が20位上がった」という。

では、手間はどうかであったかというところ。

「原稿は半角カタカナで打つので、慣れず、早く打てないので時間はかかり、特に地名は難儀した。愛媛は東予（とうよ）、中予（ちゅうよ）、南予（なんよ）と天気の子報区域が分かれるがアクセントが難しく、初回は1分の原稿に4時間近くかかった。4回目はだいぶ早くなった。」そうだ。

また、この実証実験では、働き方改革に供するものかどうかということも当然考えられた。

アナウンサーに取って代わられるという警戒感は社内に出たかという質問に三瀬氏は「仕事を乗っ取られるという反応は少ない。準備がまだ大変らしいという反応だった。」という。

また「費用対効果の点で、本来テレビのシステムである「沢村 碧」でのラジオ使用の価格として月額20万円を提示されたが、その金額ならば、ひと1人雇えるので、まだ時期尚早と判断した。」

との事だった。実験放送の結果、導入は見送られた。

まとめ

それぞれの放送局では、放送局独特の「音声合成」を放送利用する理由があった。

エフエム ワカヤマは、災害放送を少ない人数で、早朝・深夜の時間帯にも放送できる為に。NHKは、視覚障害者がテレビ放送を楽しめるように。南海放送ラジオは、新システムを放送に生かす。TBSラジオは生活の中に入ってきたAIは、「今使うとこんな感じ」を体験してほしい。だった。

現在の状況では、「ヒト」のアナウンサーと「音声合成」アナウンサーにはかなりの力の開きがあるので、すぐ取って代わるというのは難しい。だが、働き方改革として、将来第一段階は、深夜帯、早朝帯に「ヒト」がやらなくてもいいものから取って代わることは考えられる。あるいはアナウンサーが休みを取るため、決められた文言だけ音声合成に任せるという可能性はある。いずれにしても、日進月歩の「音声合成」の中で、日本語はおそらく表記でも、読みでも、一番難しい言語であり、日本語が「ヒト」と遜色なくアクセントの狂いなく流暢に音として出るのは「音声合成」の最高難度の研究だろう。

音声合成にも二つのタイプがあり、一つは、バーチャル・アナウンサー「沢村 碧」のタイプで、情報を「ヒト」が入力してそれを音にする。もう一つは、入力「ヒト」がするが、音になる前のところで機械学習がされている「Amazon Polly」、NHKの「ヨミ子」、TBSのAI犬「ドッチ君」のタイプだ。将来的には後者の機械学習タイプが、入力者の苦労がなく自動的に改良されていくので良く見える。

しかし、使用者や研究者が「ブラックボックス」といっている、機械学習の中身の状況や、その学習のための素材、資料がどこから来るものなのか分かった上でないと放送の現場で使用するのは難しい部分もある。

エフエム ワカヤマの場合は、プログラマーであった山口氏のシステムを作る力量と局の編集方針がはっきりしているため、放送ではアクセントの違和感はあるものの、地域FMとしては、少ない人数で長時間の情報発信が可能になった。台風の被害や上陸回数が多い和歌山市ならではのこ

とだ。また、そうした実績が、災害のない普段から「ナナコ」の放送を聞いている人たちが、いざという時にも聞いてくれる。と同時に、平時は、選曲機能と合わせて、ナナコにDJもやらせている。こうなるとパーソナリティーの仕事が奪うことにはならないかと考えるが、元々の人員が少ない場合は、逆に「曲選び」などという（意外に時間のかかる）手間が無くなり、他の仕事に力を割り振れることになる。

TBS ラジオの橋本プロデューサーが指摘していたのもそのあたりで、AI 選曲システムは良く出来ていて、「赤」というキーワードでの曲のリスト出しなどは簡単で、情熱⇒赤を連想する曲、歌詞に赤が含まれる曲など選曲の手間が大変省けるといい、ラジオ界を変えるかもしれないという。それをさらに「スマートスピーカー」に連動させれば、CDを放送局のライブラリーから出してくる必要がない。こうした、裏方の技術の更新が簡単になればAD（アシスタントディレクター）は他の仕事出来る。残業しなくても良くなる。そうすると、ラジオの現場の形が変わる可能性がある。また、スマートスピーカーで聞くと「○○○ラジオつけて」で、スイッチも入れない、周波数も合わせないで済むので、ラジオはスマートスピーカーで聞くべきだという。

これらの結果を考え、全くの私見であるが、今回の研究の結果、試みに今後の音声合成と放送の関わりを表で表してみた。

音声合成放送	時期	状況
第1期	2017年～	試験・実験放送、一部本放送 情報入力方式と機械学習方式
第2期	2020年～	東京オリンピックでの運用 成果出れば利用増 音声合成機械学習方式利用が増える 自動選曲等でのスタッフ的利用増加の可能性
第3期	2023年?～	スマートスピーカーの普及により音の違和感が少なくなる 音声合成DJ（ラジオの方が早いと思われる）
第4期	20〇〇年～	番組内で機械学習タイプの音声合成が出演者（テレビ・ラジオ）となりスマートスピーカー同様のやり取りをする。 一つの番組を音声合成が担当（テレビ・ラジオ）

音声合成の擬人化

エフエム ワカヤマでの『ナナコ』導入時には、違和感を持ったリスナーもいたが、毎日のニュース天気等で年月が経つと違和感が薄れる。さらに、実体がないのに名前に存在感がある。

TBS ラジオの場合も、「沢村 碧」やAI犬「ドッチ君」も実際にいないのに可愛がって、「がんばれ」などの反響も来る。スタジオでも擬人化したやり取りになる。さらに、質問された音声合成が、答えられず間が空く。それは、単にインターネットとの接続の間（ま）であったりする。しかし、ラジオリスナーは、「ドッチ君」が考えているのではないかと思う。勝手に擬人化して、悩んでるな。答えに困っているなど忖度する。こうした状態が生じるのは、やはり名前がついているからで、NHKのニュース「ヨミ子」も同じである。

だが将来的に、音声合成の「喋り」が如何に流暢になったとしても「機械の放送」であることを明記し続ける必要があると考える。特にニュースの現場で情報を入れる時は、必ず判断できる「ヒト」が行うこと。そのうちに情報を入れなくてもあらゆるものがインターネット通信でつながる

「IoT」化で、情報を入力しなくても勝手に音声合成がネット上の情報を取ってくることも可能なことになる。どこから取った情報なのか、情報入力ブラックボックス化は避けたい。ニュース編集や番組送出の責任についてもたとえ録音であっても、生放送でも、まだまだ検討する必要がある。

これは、上記の表の第3期、第4期を迎える前に検討しておくべき課題と思われる。

注. 一覧

はじめに

- (1) https://www.huffingtonpost.jp/2018/11/07/ai-anchor_a_23583263/ 動画 HUFFPOST
- (2) <https://www.youtube.com/watch?v=POjl5NYMbZU> 「沢村 碧」
<http://avataragentservice.jp> ソニー株式会社 アバターエージェントサービス 沢村 碧
<http://corp.kyodo-d.jp/> 共同通信デジタル HP

エフエム和歌山の例

- (3) <https://877.fm/table.php> FM 和歌山番組表
- (4) <https://877.fm/> エフエム和歌山HP
- (5) https://www.jcba.jp/map/area_07.html 日本コミュニティー放送協会HP
- (6) <https://aws.amazon.com/jp/polly/> Amazon Polly HP
- (7) ナナコの災害放送
<https://drive.google.com/file/d/1GCAU0Vr2ydRqqy8MnKjyWJ108AKMMzKQ/view?usp=sharing>
 西日本豪雨 2018年7月5日 エフエム ワカヤマ ナナコ声
https://drive.google.com/file/d/1AnHkjiPPRKiopXESWYOjOoBacz_t_Fl4/view?usp=sharing
 台風21号 2018年9月4日 エフエム ワカヤマ ナナコ声

TBS ラジオ好奇心家族の例

- (8) <http://www.ai-j.jp/cloud/webapi/> 日本語音声合成 AITalk (R) WebAPI 株式会社エーアイ HP
 株式会社エーアイ（本社：〒113-0024 東京都文京区西片1-15-15KDX 春日ビル10F）
- (9) <https://www.jetrun.co.jp/>
 Jetrun テクノロジー株式会社 本社〒107-0051 港区元赤坂1-7-18KIZUNA EAST ビル1階
- (10) <http://www.sockets.co.jp/>
 株式会社 ソケット 〒151-0051 東京都渋谷区千駄ヶ谷4-23-5JPR 千駄ヶ谷ビル3F

NHK の例 ニュースとロボット実況

- (11) <https://www.nhk.or.jp/ncl1-news/yomiko/> NHK ニュースヨミ子「ヨミ子の部屋」HP
- (12) 「ヨミ子」の読んだニュース
https://drive.google.com/file/d/1Xq-nnCJsC27M5CTSFr_RjkvOTqPvFmQ4/view?usp=sharing
 寝屋川二人子供殺害事件 2018年12月19日 NHK 寝屋川二人子供殺害事件判決

南海放送の例

- (13) JRN ジャパン・ラジオ・ネットワーク Japan Radio Network 1965年発足のTBSラジオ（本社・東京）をキーステーションとする民間放送ネットワーク（単独加盟4局 クロスネット30局）
 NRN 全国ラジオネットワーク NATIONAL RADIO NETWORK 1965年発足の文化放送とニッポン

放送（共に本社・東京都）をキーステーションとする民間放送ネットワーク（単独加盟10局 クロスネット30局）

(14) <https://www.rnb.co.jp/press/node/001622.php> 南海放送プレスリリース

(15) ラジコ radiko.jp 都市部高層ビル等のラジオの難聴対策や若者層のラジオ離れなどの対策として、国内のラジオ放送をインターネットで同時配信する。在京・在阪民間放送局と電通の計14社で「株式会社radiko」を2011年設立、運用開始。

(16) <https://newswitch.jp/p/15628> 日刊工業新聞 ニュースイッチ 2018年12月13日

※2 ラジコのユーザー数：ラジコでは月間ユニークユーザー数の計測方法について7月に変更し、より正確なデータを取得できるようにした。従来の計測方法の月間ユニークユーザー数は約1000万人。12年6月に初めて1000万人を突破したが、それ以来、1000万—1200万人程度で推移していた。

参考文献

「図解 これだけは知っておきたい AI ビジネス 入門」 三津村直樹著 成美堂出版 2017/9/20 発行

「俯瞰図から見える 日本型 AI ビジネスモデル」 大野治著 日刊工業新聞社 2017/12/25 初版発行

「人工知能の『最適解』と人間の選択」 NHK スペシャル取材班 NHK 出版新書 2017/11/10 初版発行

「amazon」 成毛眞著 ダイアモンド社 2018/08/08 第1刷発行

「スーパーインテリジェンス 超絶 AI と人類の運命」 ニック・ポストロム / 倉骨彰 [訳] 2018/3/6 2刷

「テレビの日本語」 加藤昌男著 岩波新書 2012/07/20 初版発行

「臨時災害放送局というメディア」 大内斎之著 青弓社 2018/10/29 初版発行

Amazon polly <https://aws.amazon.com/jp/polly/>

FM 和歌山 <https://877.fm/>

共同通信デジタル ソニー 沢村碧

<https://www.youtube.com/watch?v=POjl5NYMbzU> 2017/08/08 に公開

株式会社共同通信デジタル／ソニー株式会社による情報読み上げ動画作成システムです。文字情報として入力された原稿を音声合成エンジンによって自然な発話に変換し、さらにCGで作成されたキャラクターの表情と連動させて読み上げる情報コンテンツの提供を目的としたサービスです。

アバターエージェントサービス

南海放送 <https://www.rnb.co.jp/>

南海放送 バーチャル・アナウンサー プレスリリース

<https://www.rnb.co.jp/press/node/001622.php>

日経トレンディ ネット 2017年11月07日「AIがラジオのアシスタントに 4時間ぶっ通し出演も」

NHK ロボット実況 ニュース拡大鏡／熱戦続く平昌五輪 NHKのロボ実況が活躍

競技データ即座に音声化 (2018/2/20 05:00 日刊工業新聞 web)

NHK 技研だより 2018年3月号 NO.156

NHK 広報局 2018年3月14日 平昌オリンピック NHK 会長報告 報道資料

NHK 技研 熊野正氏講演 スポーツ番組を解説する 「音声ガイド」生成技術

NHK 技研 R&D/No.164/2018年8月号

www.nhk.or.jp/str/publica/rd/rd164/pdf/P49-55.pdf