

「キミノコトダイスキ」

「オナカヘッタ」

舌足らずの可愛い言葉を発する「ファービー」が五月末、トミーから発売された。米国では昨年登場して以来、爆発的な人気を博しているペット玩具。その日本語版だ。即日完売するおもちゃ店が続出し、国内マスコミの多くは「言語学習機能をもった人形」のヒットを報じた。

だが実際は、ファービーは言語を学習したり、認識したりはしない。最初はファービー語という意味不明の声を出し、次第に日本語を話すようになる。しかし、これは「人が音を聞かせた時間の長さに応じて、あらかじめ組み込まれている言葉を発するようになる」設計だ。人の声でなくてもいい。

「言語を学習する人形」——これはファービーに与えられた名誉ある誤解だろう。外部の音をセンサーが感知すると、ファービーは声を出す。あたかも言葉を覚えていくように見えるのだ。

話せばワイプロ文書に

「おもちゃですからね。言語はもともメモリーICに入っています」なんて言ったら、夢もないでしょう。でも、私たちはファービーが、言葉を覚えるとか、理解するとかは、一度も言っていないんですけれどねえ……」

発売元のトミーは、周囲の誤解を静かに歓迎する姿勢だ。広報担当者、やや恐縮しながらも、笑

みをこぼした。

音に反応するといえば、ソニーの大型ロボット「アイボ」も、そうだ。しかし、これも人間の言葉を聞き分けるわけではない。

人間の話し言葉を正しく聞き取って、様々な反応をする技術は「音声認識」と呼ばれる。まさに人間の命令を聞き、質問に答えるアニメの世界のロボットを連想するが、そんな技術が組み込まれた製品も、次々と出現しているのだ。音声認識技術を実用レベルまで高めた製品として、日本IBMの



ファービー(左上)やアイボ(右)は、音に反応するおもちゃだ。言葉を理解するわけではないが、愛らしくさで人を魅了する。TVゲームのピカチュウ(左下)なら、人の言葉がわかる。マイクに向かって命令すると、その言葉に応じた動作をしてくれる

人の声を聞く スーパー機械

音に反応する楽しい新製品と音声認識技術

人間の声に反応するおもちゃや機械が話題を呼んでいる。すべてが人の言葉を判別しているわけではないようだが、人の言葉に応じて機械が動くのは、考えるだけで楽しい。

編集部 伊藤隆太郎 写真 稲垣徳文

会場で、その日の新聞を、ピアボイスに向かって読み上げる。見る見るうちに、記事そっくりの文章が、パソコン画面に現れるのだ。「いつも、数行を読むあたりでデモ会場がざわざわして、『おぉ』という驚きに変わるんですよ。SF映画みたいで信じられないのでしようね」

信じてもらうためにも、その日の新聞を使うのだと、岡部さんはいう。誇らしげな表情だ。

車を運転しながら対話

なぜ、ほかの製品を引き離すような完成度を持つ開発が可能だったのか。ここはまず、他社の技術者に語ってもらおう。大手メーカーの音声認識担当者はいこう。

「認識の理論自体には、大差ないはずですが、でも、それを実用化させるまでの蓄積が違う。人それぞれの話し言葉がもっている特徴を、膨大なデータに置き換えて、積み上げた。いわば、費やされた研究時間の差でしょう」

その「膨大な研究」を積み重ねてきたのが、米国にあるIBMワトソン研究所だ。人間の声を認識する際に、独特の確率的な処理を施す方法が、二十年以上も追究されてきた。

日本語版の音声技術プログラム担当、金子宏さんも一時期、ワトソン研にいた。

「人間の声を文章に置き換える理論は、英語も日本語も同じです。だから、ワトソン研の成果が、そ

音声認識ソフトウェア「ピアボイス」が知られている。マイクに向かって話した言葉が、そのままワイプロ文章になるのだ。パソコン自体を声で操作することもできる。使い方に慣れれば、音声認識率は九〇%を超すという。昨年夏に発売され、年末までに三十万本を売った。パソコンソフトとしては大ヒット。数々の賞も受けた。

ソフトウェア事業部の責任者、岡部春樹さんは時折、企業や団体から実演を頼まれて、出かける。



のまま日本語に応用できた」

と金子さんはいう。

もちろん、日本語化にあたっては、たくさん日本人の音声データが蓄えられた。被験者は千人にのぼる。なかには、最大のライバル会社ともいわれるNECの系列社員までいた。五年前に偶然、音声データの提供者になったのだ。

「IBMの東京基礎研究所で、いろんな文章を、決められた話し方と区切り方で読みあげる作業を繰り返した。一般公募はできないからと、ある特別な方法で選ばれたら、それが自分だった」

と、苦笑いしながら振り返る。熾烈な開発競争の舞台裏では、そんな皮肉な偶然もあったのだ。

音声認識の完成度を激しく競い合っている代表的な分野として、ほかには、自動車内で道案内をするカーナビがある。運転中などでも目的地を入力しやすいように、音声認識を利用する。

六月四日、京都市で開かれた人工知能学会の「言語・音声理解と対話処理研究会」では、東芝の研究開発センター研究主務、河野恭之さんが、新しいカーナビ試作機を発表した。

高速道路を走っている場面で、人間とカーナビの「会話」は、こんな具合だ。

「目的地までどれくらい？」
「およそ二十キロ、三十分程度かかります」
「そこまでの最後のコンビニは？」

「〇〇コンビニ××町店です。しかし、駐車場はありません」
かなりくだけた話し言葉を、カーナビが正しく認識して、二秒程度で返答してくる。その一方で、こんな会話もできるのだ。

「最初に左折する交差点から最後

に右折する交差点まではどれくらいですか？」

人間が一気に問いかけると、カーナビもすぐに返答した。すごい。だが、実はこれは、むしろ機械にとっては簡単な質問なのだ。

「現実には人間が、こんな『整然とした質問』なんて、できませんよね。途中でつかえたり、言い間違えたりする。その処理が難しいし、今後の開発課題でもある」

声で遊べるTVゲーム

カーナビでは、音声入力ワープロのような「パーセントの認識率」を目指すよりも、むしろ重要な単語をきちんと選び出すことのほうが大切になる。また、車の運転席は風やエンジン音などにあふれているから、雑音から人の声を抜き出す技術開発も、ワープロと

は違った課題だ。

河野さんは、あらゆる音声関連の技術動向を注視している。

ファービーも真っ先に入手した。米国で発売されると、研究仲間といっしょに取り寄せて、試用した。なにしろ最初は、「秘かに録音装置がついている」などと噂が飛び交い、アメリカ海軍が持ち込みを禁止する騒動まであったのだ。結局、音声認識などはなかった。

だが、そんなおもちゃの世界にも最近、音声認識を搭載した製品が登場している。昨年発売されたTVゲーム「ピカチュウげんきでちゅう」は、家庭用ゲームで世界初の音声認識システムだ。

声で命令をしながら、ゲームを楽しむ。八十万本のヒット商品になったが、意外にも、認識の精度自体は決して高くない。だが、河

野さんはこう評価する。「音声の認識間違いが、あまり問題にならないような製品設計や利用者層の設定になっている。そこが見事だ」

つまり、言うことを正しく聞いてくれなくても、楽しいのだ。東京・板橋の小学生、森田真菜美ちゃん(10)、運香ちゃん(7)姉妹はこのゲームが大好き。四十本近く持っているTVゲームの中で、一番のお気に入りという。

電源を入れると、TV画面で、ピカチュウが寝ていた。真菜美ちゃんがマイクから呼びかける。

「ピカチュウ、起きて」

ピカチュウが目を覚まし、笑いながら近づいてきた。

魚釣りをしている場面では、

「引っ張れ」

と応援すると、ピカチュウががんばって釣りざおを引く。魔法をかけたときは、

「十万ポルト」「電気ショック」かみなり」

などと命令する。

キーワードは「楽しさ」

だが、言葉が理解できないとき、ピカチュウは頭をかいて、顔を赤くする。その動作がなかなか面白い。

「ピカチュウは、言葉を聞き間違えることも多いけど、とにかく何かの反応してくれるし、しぐさが可愛い。だから、とても楽しい」と真菜美ちゃん。ピカチュウが転んだり、照れたりする場面では、



ピアボイスを開発した日本IBM(右上)では、パソコンに入力する日常業務も音声認識で進めている。日立の音声認識通訳機(右下)も、スイッチ類は数個しかなく、主な操作はすべて音声だ。ミサワホームの住宅(左)は、「暗いよ」と呼びかければ、明かりがともる

子供たちは大笑いになった。

音声認識という技術を見渡すと、中核となるキーワードの一つは、「楽しさ」ではないだろうか。

「自分で言うのもおこがましいのですが、やっぱり楽しいんですよ。電池がなくなるまで、遊んでしまいました」

日立グループは五月、携帯型の音声認識通訳機を試作、発表した。

開発の中心となった日立超LSIシステムズの北爪吉明主任技師は、そうやって笑う。

大きさは電卓程度で、胸ポケット

つまり、本格的な外国語翻訳機ではなく、手軽な英会話学習機を目標にしたのだ。

「おやすみ」の声で消灯

音声認識の部分を担当した日立製作所中央研究所の大淵康成さんという。

「間違った日本語を話したら、厳密に『間違った英文』へと訳すような、そんな翻訳機が必要でしょう。それよりも、ユーザーが気兼ねなく使ってもらえるものを、機械が幅広く取り扱ってあげる優しさ。それこそが、音声認識技術でしよう」

人間の側が、どんな日本語を出してよいか迷わなくていいのだ。とにかく思いついた言葉を入れれば、機械が次々と、様々な英文を話してくれる。それが、楽しさに結びついている。

ミサワホームも住宅機器を声で制御するシステムを開発した。

「開けゴマ! でドアが開くと面白いですね」

きつかけは、そんな社長の一言。やはり、スタートは「遊び心」なのだ。

同社総合研究所の酒井靖夫エレクトロニクス研究室長は、この一言を受けて、数年がかりで実現にこぎ着ける。

「シャッターを下げて」

マイクにそう命じると、静かにシャッターが下りてくる。認識できる単語数は約五百語。工夫をしたのは、できるだけ「機械らしさ」

を隠すことだった。命令は、こんな感じでもいい。「あついなあ」

すると、エアコンの電源が入り、室温を下げられる。「おやすみ」

と命令すれば、まずTVやエアコンが切れ、さらに十秒後に、室内の明かりが消える。「住宅メーカーとして蓄積した人間の生活パターンについての研究成果を生かしているのです」

と、酒井さんはいう。音声認識技術を取り入れた製品は、一九八〇年ごろから、目に見える形を現す。日本ではNNTTが八一年、電話で銀行の残高照会などができる音声認識システム「ANSER」を開発している。

最初は、人間の声と、機械が知っている音声とのマッチング具合を照合することに、開発の重点が置かれた。その後、九〇年ごろから「確率モデル」が導入され、人間の言葉が持つ多様さや曖昧さを、うまく修正できるようにしてきた。音声認識技術はここ十年間で大きな進歩を見ている。

NNTTサイバースペース研究所の山森和彦メディア処理プロジェクト担当部長はいう。

「つまり、これまでの十年とは、音声認識の精度を追求する十年だった。これからは、精度を高めた音声認識が、実際に製品に搭載され、人間にとっての『使い心地』や『扱いやすさ』を高めることが、追求される時代となるだろう」