

[最近のトピックス]

音声合成ソフトによる「自分の声」での会話システム

廣瀬 知二

伊東歯科口腔病院

頭頸部がん治療のための喉頭摘出術や、ALS（筋萎縮性側索硬化症）の進行などで声を失う人は年間1万人程度いるといわれている。声を失うことは、身体的、精神的負担が大きく、QOL低下の大きな要因となる。したがって声をとり戻すことは治療意欲を継続していくうえでも重要である。

現在、喉頭摘出術後には代用音声として、食道発声、電気喉頭、シャント発声の3つの方法がある（福島, 2017）。食道発声法は空気を食道内に摂取し、その空気を逆流させて食道上部を振動する行う方法で、習得するまでには数カ月～2年程度の練習が必要である。電気喉頭は振動する機械を首に当て、咽頭粘膜を震わせて音を出す方法であるが音声の抑揚性が乏しい。シャント発声は、気管と食道を手術でつなぎ、肺からの空気を食道に流入させ食道粘膜を振動する方法で、他の代用音声に比べて、本来の喉頭発声に一番近い方法である。しかし手術が必要なこと、誤嚥性肺炎のリスクを伴うなどの短所がある。いずれの発声法も患者が以前有していた音声の再現とはいえ、またALSのような進行性疾患では適応することが困難である。

一方、代用音声以外のコミュニケーション方法として、音声を自動的に生成する音声合成技術が応用されている。音声合成ソフトは種類が多く、フリーソフトで入手することも可能である。しかしながら、50音の単音を収録し、その50音を1文字ずつつなぎ合わせて合成音をつくる方法では、機械的な音となってしまう、違和感がなく会話ができるとはいえない。ボイスター[®]（ヒューマンテクノシステム）は声を失う前に利用者本人の声を録音し、データベース化することにより「自分の声」に近い音での会話を可能とするシステムである。

まず、利用者の日常会話や仕事の内容などを聞き取りし、どういう内容の言葉を発声する機会が多いかを調べる。その上で、定型文に加え特徴的な言い回しなどを加えた文章を朗読して録音する。その後、得られた長い音声ファイルを短い複数のクリップに分割する。次に実際の音声データに対して、その聴取結果や波形/スペクトルの視察結果に基づいて、発話内容に則した音素ラベル付与と各音素間の境界位置を特定するラベリングと呼ばれる作業を行う（図1）（鷹見ら, 2001）。そして、このラベリングしたデータをもとにデータベースを構築して音声を合成する。これでパソコンのキーボードに文章を入力すると自由に話すことができる。

現在、外出時の利便性を改善するため、スマートフォン向けのシステムが開発されつつあり、実用化が待たれる（図2）。

文献

福島啓文. 喉頭摘出後の代用音声. 日本臨牀 75 (増刊号2) : 533-538, 2017.

鷹見淳一, 加藤喜永, 呂 彬, 佐藤奈穂子. 半自動音声ラベリングシステムの開発. Ricoh Technical Report 27 : 43-51, 2001.

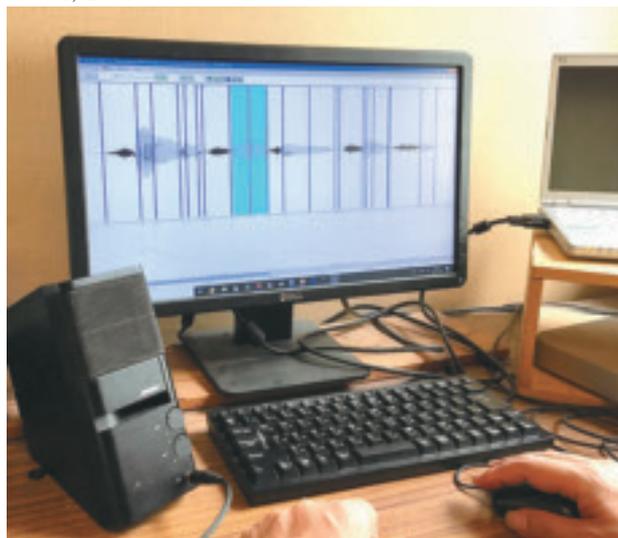


図1 音声データベースを構築する際に最も労力を要する作業はラベリングである。発話内容に適した音素ラベルの付与とその境界位置を決定する。



図2 タブレット端末でも使用は可能。スマートフォンでのシステムは試作段階である。実用化により出先での会話もよりスムーズに行えることが期待されている。