サイエンスBOX

探る　学ぶ　知る　聞く

騒音との違いAI学習

「音声認識システム」

叫び判別 防犯に一役

街のどこかで助けを求める叫び声。たくさんの騒音の中から、その叫び声だけをしっかりと聞き分ける新しい音声認識システムの開発に、立命館大の研究チームが取り組んでいる。街頭の防犯カメラなどに活用すれば、大きな効果がありそうだ。研究者を訪ね、その性能を体験した。

（松田俊輔）

研究チームの代表は、立命館大情報理工学部の福森隆寛助教（音声・音響学）。滋賀県草津市にあるキャンパスを訪れると、さっそく実験場所へと案内してくれた。

室内にスタンド型のマイクが1本立っている。そばにはノートパソコンと大型画面。このマイクで音を拾い、叫び声なのか、ただの大きな音なのか、パソコンの人工知能（AI）が識別するのだそうだ。

「まずは叫んでみてください」。指示されるがままに声を張り上げた。「わーっ」。1秒後、画面に「叫び声　81・5％」と表示された。これは、「今の音が叫び声である可能性は81・5％ですよ」という意味らしい。

「次はただの大声で」。先ほどの叫び声と同じぐらい大きな声で、今度は「助けてー」と発してみた。表示は「平静音声（異常なし）」と出た。「叫び声ではなく、普通の声ですよ」ということだ。

少し声を張り上げ、もう一度「助けてー」と発すると、「叫び声　65％」と表示が変わった。

AIは何をもとに「叫び声」と認識しているのだろうか。

■□■

福森助教によると、AIは、喉にある「仮声帯」という部位の動きと、口や喉の形の二つを指標にして、叫び声かどうかを判断しているのだという。

通常、声を出す時は声帯が大きく振動する。一方、叫ぶ時は喉を絞るように使うため、声帯だけでなく、その上にある仮声帯も大きく振動する。

仮声帯の動きの大きさをグラフにすると、叫び声では鋭い山のような形が現れる。また、口や喉の形に着目して、周波数と音の強さをグラフにすると、叫び声の場合はギザギザした形になる。

研究チームは、10～20代の50人に、通常の声と叫び声をそれぞれ10回ずつ発してもらい、二つの指標をもとにグラフを作成した。そして「ディープ・ラーニング（深層学習）」の技術でAIにグラフの形を学習させ、今回の判別システムを作ったという。

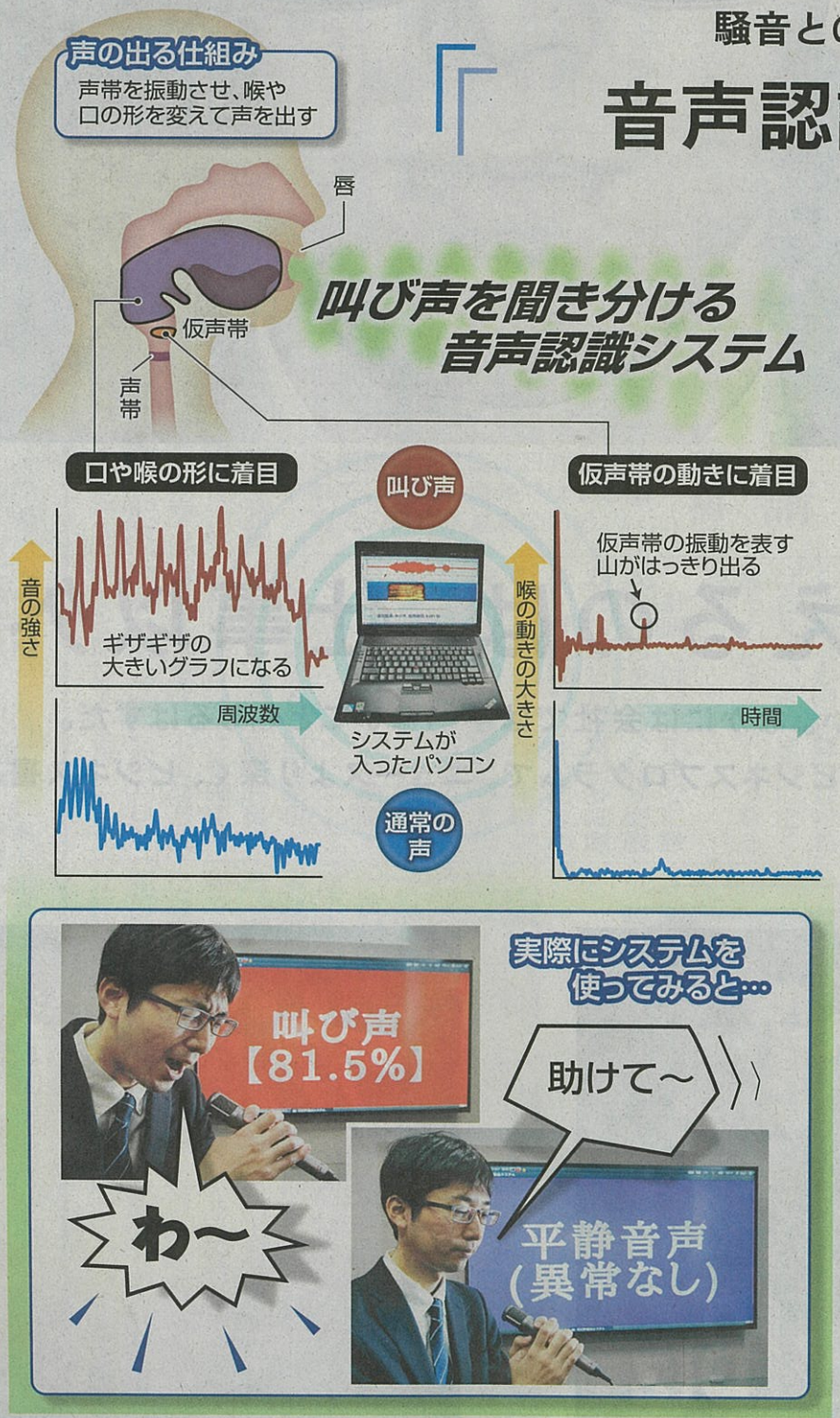
つまり、先ほどの「助けてー」では、喉を絞るように使わなかったので、グラフの形に叫び声としての特徴が表れなかったということのようだ。

■□■

このシステムが完成した暁に想定されるのが、街頭の防犯カメラへの搭載だ。犯罪被害者が発した叫び声を、街の雑音の中から聞き分け、素早くカメラを向けることもできるようになるという。

また、防犯ブザーやスマートフォンの防犯のアプリに搭載すれば、叫び声をキャッチした時に自動的にブザー音を鳴らすということもできるだろう。

福森助教は「技術はほぼ完成している。協力できる企業などを探し、実用化を目指したい」と話す。実用化されれば、街頭犯罪の抑止に貢献するに違いない。



口の形、仮声帯の動き 着目

人間が声を出す場合、まずは声帯を振動させて音を発生させる。さらに音の通り道となる口や喉などの形を変えることによって、様々な言葉を話している。

通常の声を出す場合、声帯が1秒間に振動する回数は男性で約120回、女性で約250回といわれている。叫び声や笑い声など高い音を出す場合には、振動回数が増える。

また、大きな声を出すには口を大きく開ける必要がある。叫ぶ時はこれに加え喉を絞る。叫んでいる時の喉の形は、通常より部分的に狭くなっているという。

叫び声を周波数や強さなどに着目して解析していくと、こうした声帯の振動回数や喉、口などの形の変化を受けて波形などに特徴が生じる。叫び声を聞き分ける研究では、こういった特徴をうまく組み合わせることで、識別の精度を高めている。