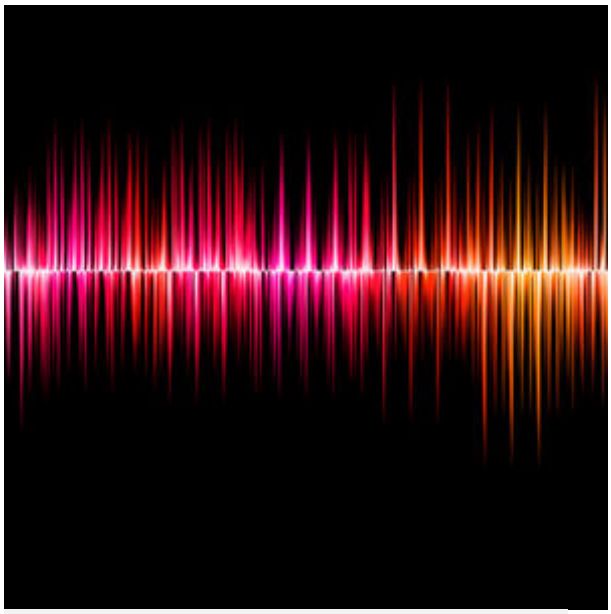
Hidden Voice Commands

Spracherkennungssoftware als Angriffsziel

09.03.17 | Autor / Redakteur: Dr. Bernd Schöne / Peter Schmitz



Künstlich erzeugte Klänge, die digitale Assistenten wie Alexa oder Google Now manipulieren, gehörte bislang nicht zu den etablierten Angriffsvektoren von Cyberkriminellen. Das könnte sich ändern. (Bild: Pixabay)

Spracheingabe und „Always-on“ liegt voll im Trend. Amazons Echo wurde ein Bestseller. Bei mobilen Geräten haben sich die Spracherkennung-Features Cortana (Microsoft) und Siri (Apple) durchgesetzt. Jetzt ist es Forschern gelungen, die Mensch-Maschine-Schnittstelle zu korrumpieren und Malware zu verteilen.

Für viele Nutzer von Computern, Smartphones oder Autos ist die eigene Stimme mittlerweile die primäre Eingabemethode. Alle modernen Sprachdienste funktionieren ähnlich. Neuronale Netze oder Algorithmen auf der Basis von Markow-Ketten setzen die menschliche Stimme in Computerbefehle um.

Auf der Sicherheitskonferenz IT-Defense 2017 in Berlin stellte der amerikanische Doktorand Tavish Vaidya von der Georgetown University nun mehrere Wege vor, die Mensch-Maschine-Schnittstelle zu korrumpieren. Das fatale dabei ist, dass nicht nur der Computer an der Nase herumgeführt wird, sondern auch der Mensch. Er hat kaum eine Chance, die Form von Malware zu erkennen. Zu allem Überfluss ist die Erfolgswahrscheinlichkeit recht hoch. Falsche Befehle erkennt der Computer präziser, als die vom Nutzer gesprochenen.

Sprachanalyse-Software ausgetrickst

Malware kann heute über die unterschiedlichsten Wege verbreitet werden. USB-Sticks und E-Mail sind nur die bekanntesten. Aber auch in Tönen lassen sich versteckte Informationen unterbringen. Künstlich erzeugte Klänge, die Spracherkennungssoftware manipulieren, gehörten bislang nicht zu den etablierten Angriffsvektoren. Das könnte sich ändern.

Forscher der amerikanischen Universitäten Berkeley und Georgetown haben sich des Themas im letzten Jahr wissenschaftlich angenommen (pdf). Sie analysierten die heute verwendeten Sprachanalyse-Software und versuchten sie auszutricksen. Sie fragten sich: Welchen Input brauchen diese Algorithmen, um einen beliebigen Output zu erzeugen?

Es gelang ihnen, für den Nutzer fast unhörbare Befehle abzusetzen, und so mit Schadcode infizierte Webseiten aufzurufen, oder gezielt Fehlfunktionen in angeschlossenen Geräten hervorzurufen. Einzige Voraussetzung für den Angriff ist ein für die Mikrofone des Computers gut hörbares Audiosignal, möglichst ohne Echo und störende Geräusche. Die unerwünschten Befehle könnte also über Radio, Fernsehen oder Lautsprecheranlagen übermittelt werden, ebenso eignen sich Freisprechanlagen von Handys. Auch in Tonträger könnte man sie einarbeiten, quasi als ungebetene Zugabe zur Musik.

Zu erkennen sind die Befehle nur schwer. Die Forscher konnten demonstrieren, dass sich eine komplexe Befehlsfolge wie: Gehe zur Webseite "evil.com" in einer Mischung aus Knacksen und Krächzen unterbringen lässt. Möglich macht dies die exakte Kenntnis der verwendeten Algorithmen. "Mensch und Computer verstehen Sprache total unterschiedlich", erläutert Tavish Vaidya. Genau hier setzen die Angreifer an. Sie wissen, dass die von der Elektronik aufgenommenen Wellenzüge zunächst in viele, sich überlappenden Teile zerlegt werden und anschließend als Kette von Messwerten zur eigentlichen Spracherkennung gelangen.

„Eine Herstellerspezifische Software analysiert die Wellenzüge, die das Mikrofon liefert“, erläutert Tavish Vaidya. Doch die Wissenschaftler bemerkten, dass nur ein Bruchteil der anfallenden Werte auch zur Analyse verwendet werden. Die Hersteller gehen davon aus, dass stets ein Mensch spricht, und wie die menschliche Stimme funktioniert, ist ihnen gut bekannt. „Von 1000 Messwerten, werden nur 50 verwendet, um daraus die gesprochenen Worte zu rekonstruieren“ erläutert Tavish Vaidya, „der Clou unserer Angriffe war, dass wir den Analysewerkzeugen nur diese 50 Werte lieferten und die anderen gar nicht erst erzeugten.“ So ergeben sich Klänge, die für Menschen verstörend fremd und meist völlig unverständlich klingen, von der Maschine aber mit höchster Wahrscheinlichkeit im Sinne des Angreifers interpretiert wurden. Die Erkennungsquote für den solchermaßen „mundgerecht“ zubereiteten Töne war mit 82 Prozent sogar höher als bei realen Eingaben des Besitzers, von denen nur 74 Prozent beim ersten Mal korrekt erkannt wurden. Der Mensch hat kaum eine Chance, die manipulierten Klänge als solche wahrzunehmen. Von 377 Versuchen im Labor mit Testpersonen verlief nur ein einziger erfolgreich.

Doch es gibt bei dieser Form der Attacke eine hohe Hürde, die der Angreifer zu überwinden hat. „Er muss die Funktionsweise des betreffenden Gerätes und die verwendete Software zur Spracherkennung genau kennen“, erklärt Tavish Vaidya. Mit

diesem Wissen ist es möglich, die Spracherkennungsfunktion umzudrehen und so rückwärts jene Wellenzüge zu berechnen, die zur gewünschten Texteingabe führen.

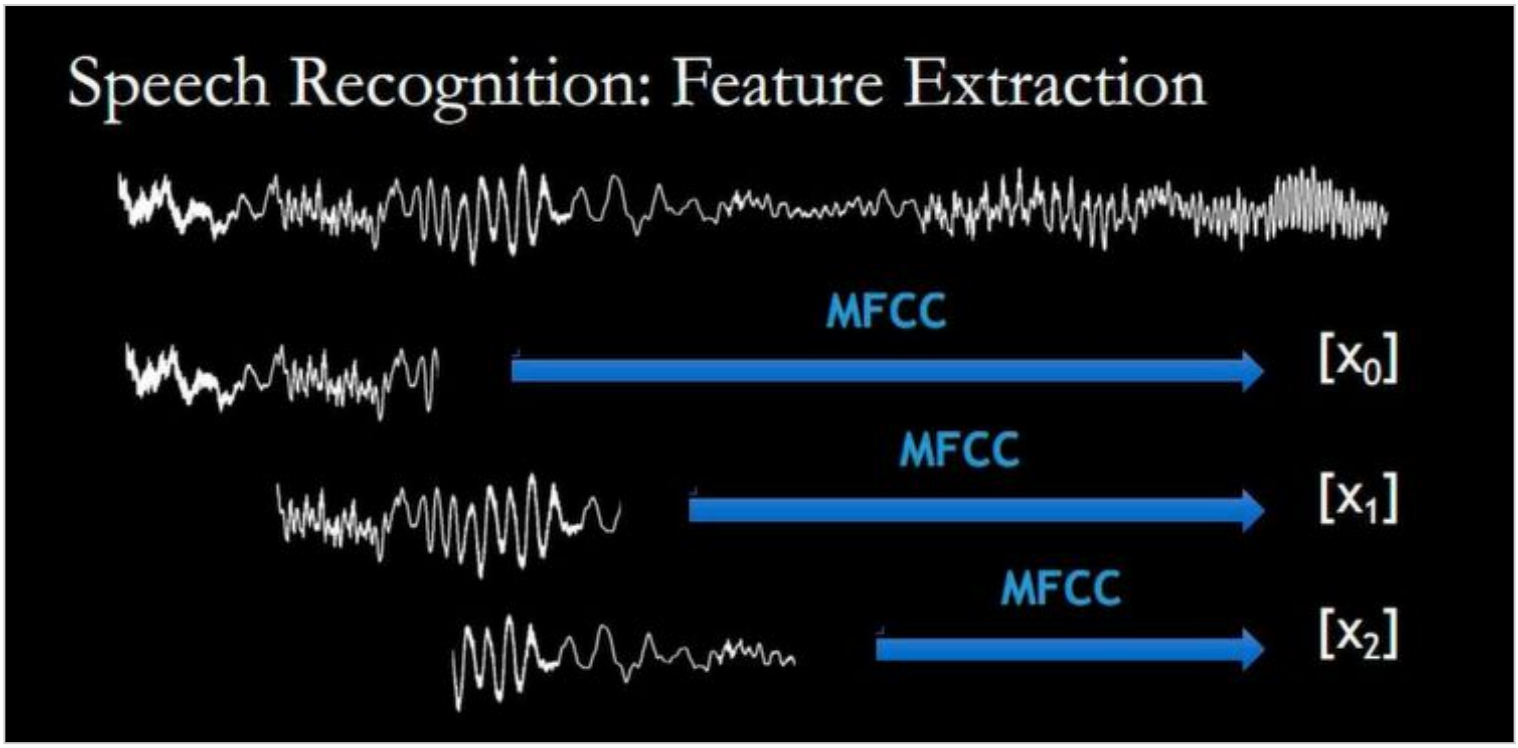
Wenn der Angreifer die Algorithmen zur Spracherkennung nicht kennt, steht ihm ein zweiter Weg offen. „Wir haben auch Black Box Angriffe durchgeführt und sind damit ebenfalls zum Ziel gelangt“, erläutert Tavish Vaidya. Bei dieser Methode probiert der Angreifer einfach so lange bestimmte Wellenzüge aus, bis die Spracherkennung die gewünschten Texte liefert. Auch hier ist die Trefferwahrscheinlichkeit hoch, wenngleich einige Prozentpunkte geringer wie bei der ersten Variante. In jedem Fall muss der Angreifer zumindest Hersteller und Typ des Opfersystems kennen, um die versteckten Befehle übermitteln zu können. In der Praxis werden sich Angreifer also auf weit verbreitete Systeme konzentrieren. Wie der Black Box Angriff funktioniert, zeigt das folgende Video:

Unüberschaubares Risiko

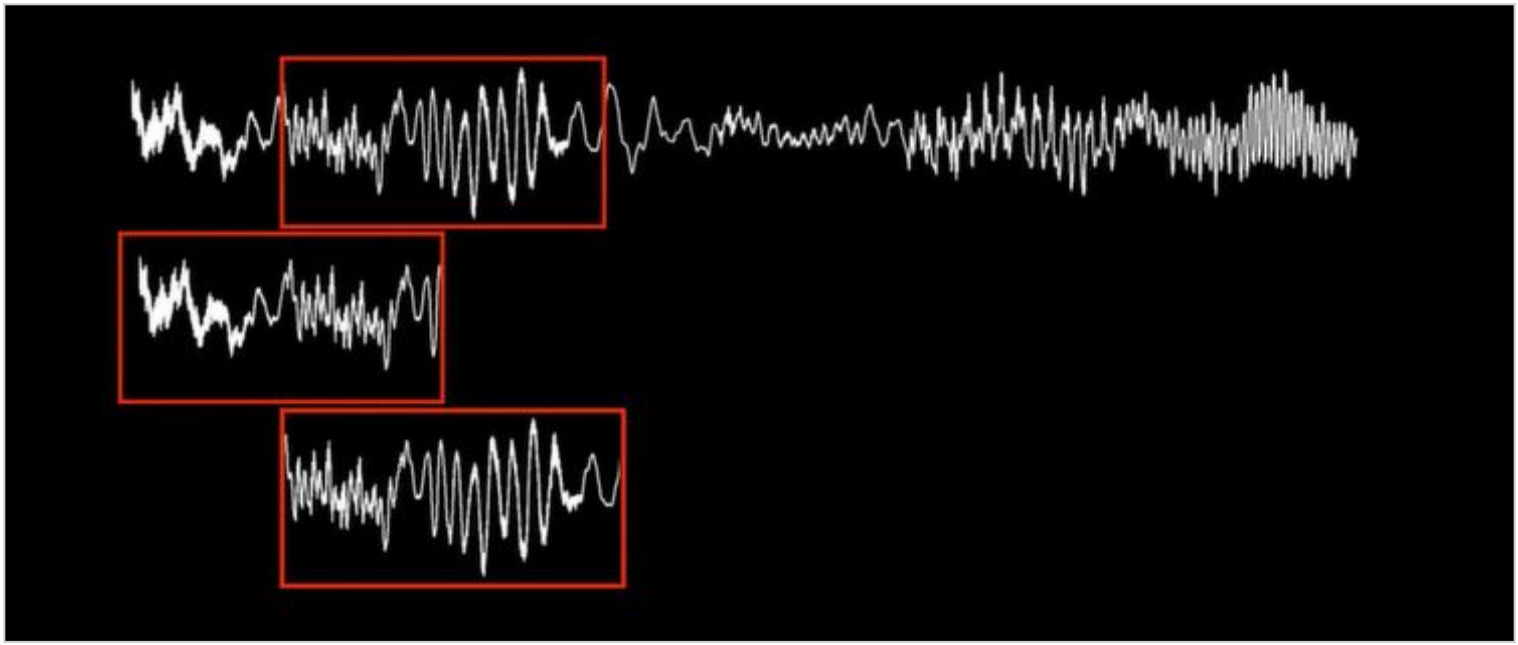
Trotzdem ist das Risiko nicht zu unterschätzen, warnen Experten, vor allem nicht in Hinblick auf die Möglichkeit, beliebige Webseiten aufrufen zu können, und von dort fremde Dienste aufzurufen. Auch solche, die Geld kosten oder über eingeschleuste Schadprogramme Passwörter und Verschlüsselungstrojaner starten.

Inzwischen verfügen auch Autos über Spracheingabesysteme. Ein starker Sender könnte sehr schnell das Wunschprogramm des Fahrers übertönen, und fremde Musik ins Fahrzeug bringen. Der Schadcode wäre dann in kleinen Störungen innerhalb der Musik verborgen, die während der Fahrt völlig unverdächtig sind. Auch eine direkte Schädigung ist denkbar, dazu müsste der Angreifer nur im Namen und auf Rechnung des Opfers bei einem Onlinehändler einkaufen.

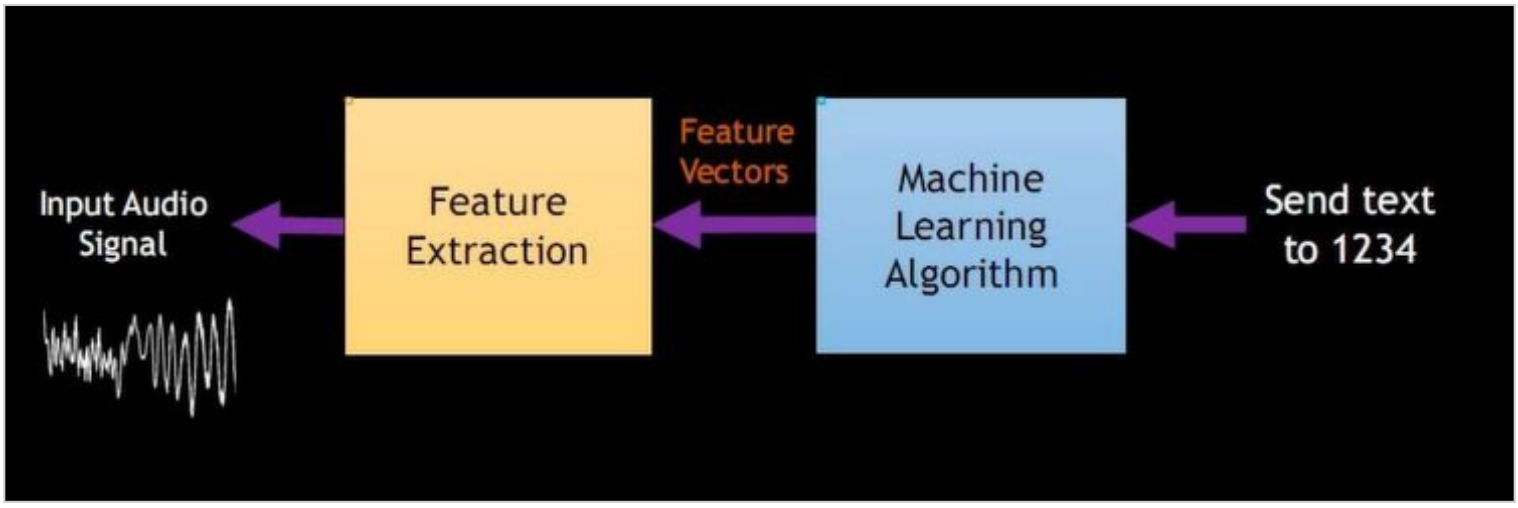
Auch normale Radio- und Fernsehprogramme eignen sich theoretisch. Angreifer könnten einen Werbespot mit Malware anreichern. Etwas Ähnliches ist Anfang des Jahres bereits geschehen, allerdings eher aus Zufall. Anfang 2017 gingen zahlreiche Amazon-Echo-Geräte eigenständig und unerwünscht auf Einkaufstour. Sie orderten Puppenhäuser, obwohl die keiner bestellen wollte. Auslöser war ein Beitrag eines kalifornischen Senders in der News-Sendung CW6. Dort tauchte der Satz „Alexa hat mir ein Puppenhaus gekauft“, auf. Auf Englisch hört sich der Satz „Alexa ordered me a dollhouse“ an wie die Aufforderung „Alexa, order me a dollhouse!“ Auf das Schlüsselwort „Alexa“ reagierten die stets lauschbereiten Echo Geräte. Da zudem Standardmäßig Online-Einkäufe per Sprachbefehl zugelassen sind, stand der erfolgreichen Bestellung nichts mehr im Wege. Ein Fall, der Spitzbuben auf fiese Gedanken bringen könnte.



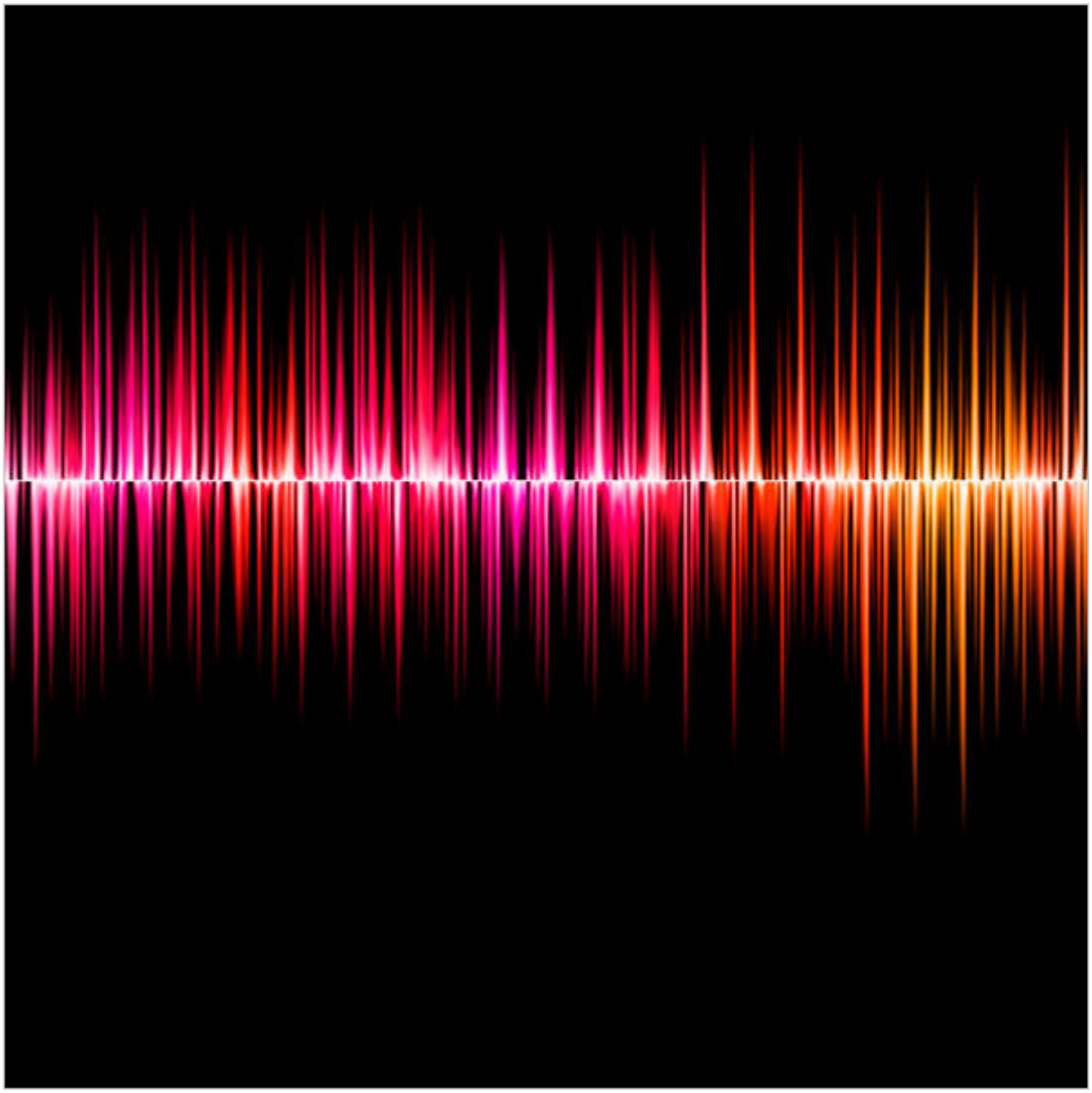
Die von der Elektronik aufgenommenen Wellenzüge werden zunächst in viele, sich überlappende Teile zerlegt und anschließend als Kette von Messwerten zur eigentlichen Spracherkennung weitergereicht. (Tavish Vaidya)



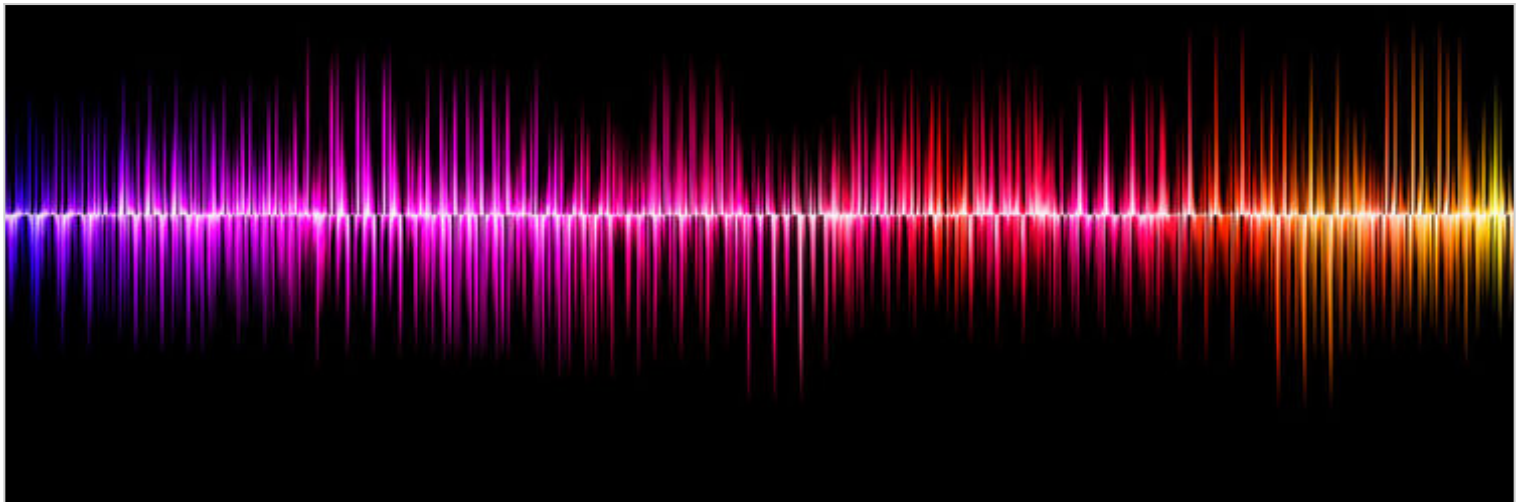
Die Wissenschaftler bemerkten, dass nur ein Bruchteil der anfallenden Werte auch zur Analyse verwendet werden. Von 1000 Messwerten, werden nur 50 verwendet, um daraus die gesprochenen Worte zu rekonstruieren. (Tavish Vaidya)



Die Erkennungsquote für die als Angriff erzeugten Tonsequenzen war mit 82 Prozent sogar höher als bei realen Eingaben des Besitzers. (Tavish Vaidya)



Künstlich erzeugte Klänge, die digitale Assistenten wie Alexa oder Google Now manipulieren, gehörte bislang nicht zu den etablierten Angriffsvektoren von Cyberkriminellen. Das könnte sich ändern. (Pixabay)



Künstlich erzeugte Klänge, die digitale Assistenten wie Alexa oder Google Now manipulieren, gehörte bislang nicht zu den etablierten Angriffsvektoren von Cyberkriminellen. Das könnte sich ändern. (Pixabay)