

“

Tak jako má každý člověk jedinečnou DNA, jedinečný rukopis nebo nezaměnitelné otisky prstů, má také jedinečné mluvící ústrojí, ve kterém vzniká hlas.

a druhá je usvědčovat a honit kriminálníky, zloděje nebo atentátníky. Máme třeba projekt, který je podporován českým ministerstvem vnitra, a my se snažíme o to, aby jeho složky měly k dispozici to nejlepší, co vymyslíme. My s nimi spolupracujeme, ovšem jen do té míry, že jim dodáváme naše věci a získáváme od nich poznatky.

Která část české policie má o váš systém zájem?
Řekl bych vám to rád, ale jsem vázán mlčenlivostí. Nicméně máme zprávy, že naše technologie je velmi úspěšná. Mimochodem, zájem o ni mají vojenské složky i v zahraničí, například ve Spojených státech i v Izraeli. Slouží jim především k tomu, aby se rychle dokázaly prohrabat nepředstavitelně velkým množstvím nejrůznějších audionahrávek pořízených z monitorování podezřelých osob, které mluví nejrůznějšími jazyky. Nikde na světě není tolik lidí s jazykovými znalostmi, kteří by dovedli v krátkém čase rychle a přesně vyhodnotit klasickým poslechem tisíce nahrávek.

Představte si, že máte za úkol určit z tisíců odposlechů dva tři hovory, ve kterých se objevil člověk, který je spojován s velkou bankovní loupeží. Nehledal byste jehlu v kupce sena?

Ano i ne. Úplně na začátku bych se samozřejmě mohl zaměřit na nějakou technickou informaci, například na číslo mobilního telefonu nebo SIM karty, které síť automaticky přenáší. Tohle by byla pro mě základní informace, ale jen podpůrná. Pak bych měl další možnost, a to, že bych si najal, řekněme, padesát lidí a řekl jim: Poslechněte si každý záznamy tisícovky různě dlouhých hovorů a pak mi řekněte, kde jste toho lumpa zaznamenali. Buďte si jistý, že by jim to trvalo hodně dlouho a spolehlivost by byla velmi pofiderní. Při využití našeho systému to bude trvat řádově stokrát rychleji. Například dvou- tříminutová nahrávka jen několik vteřin. Čas by se ještě podstatně zkrátil ve chvíli, kdybychom všechny záznamy prohnali serverovou farmou, tedy systémem propojených výkonných počítačů.

Před necelými dvěma týdny se Usáma bin Ládín opět nechal slyšet ve třináctiminutové nahrávce; vyzývá v ní své přívržence, aby se zaměřili na teroristické akce v Evropě. Jak se zjišťuje pravost nahrávky, tedy i to, zda nejde o podvrh?
Analytici ve Spojených státech, kteří se dostanou k záznamu, využívají všechny možné technologie pro verifikaci mluvčího, aby nakonec řekli, zda to je, či není on. Ale oni se nespokojí jen s analýzou hlasu. To, co udělá software se zvukovým záznamem, se děje i s videem, pokud je k dispozici obrazový záznam. Nejdůležitější je ale skupina lidí, která nedělá nic jiného, než že ušima poslouchá hlas a očima sleduje gesta, tvář a pohyby Usámy bin Ládína. Znají ho velmi dobře. Jsou to skvělí someliéři zvuku a obrazu. Právě ti mají to nejdůležitější slovo v konečném verdiktu, zda to je, či není Usáma.

Má elektronická analýza lidské řeči šanci uplatnit se i v běžném životě?

Navrhli jsme několik scénářů, kde si myslíme, že by náš systém mohl velmi úspěšně fungovat. Například v call-centrech obchodních firem nebo finančních institucí, které obsluhují zákazníky po telefonu. Představte si, že si potřebujete zvýšit limit na výběr ze své kreditní karty nebo si přes obchodníka chcete koupit velký balík cenných papírů na burze. Standardně to je tak, že mu zavoláte a identifikujete se tím, že namačkáte na telefonu číselný kód nebo obchodníkovi ústně sdělíte například první, páté a osmé číslo či písmeno svého hesla. Jenže! Když se k číselnému kódu dostane někdo cizí, a to dnes není žádný velký problém, pak si dovedete představit, kam mohou vaše peníze odtéct. No a my si představujeme, že třeba bezpečnost nejrůznějších převodů se může pojistit vašim hlasem, podobně jako se pojišťuje vstup do sejfu otiskem prstu, dlaně nebo oční rohovkou.

Jak byste si to představoval konkrétně?

Software by ověřil, zda ten, kdo volá a žádá nějakou službu, je skutečným majitelem účtu či portfolia cenných papírů. Co by tomu muselo předcházet? Při založení účtu nebo při registraci u brokera by zájemce musel namluvit minutovou nahrávku svého hlasu. Když pak bude volat s žádostí o službu, systém propojený z telefonu do počítače bude schopen ve zlomku vteřiny vyhodnotit, jestli je to skutečný majitel účtu, nebo podvodník. Pro ověření stačí tři čtyři věty. Verifikace hlasu není stoprocentní, proto kdyby operátor viděl, že skóre věrohodnosti je velmi nízké, mohl by vás potrápít dalšími doplňujícími otázkami. Anebo by vás nakonec donutil přijít k přepážce.

Pozná váš počítačový software, že někdo úmyslně mluví jiným hlasem, mění dikci nebo záměrně špatně artikuluje?

Popravdě řečeno, nepozná, protože bude zkreslení nebo karikování hlasu považovat za normální projev mluvčího. My se ale snažíme o to, aby se s těmito věcmi dokázal vyrovnat. Můžeme ho to učít tak, že vezmeme stovku lidí a hlas každého z nich nahrajeme v různých podmínkách. Jednou bude mluvit v klidu, podruhé unavený, potřetí opilý a v posledním případě nebude mluvit česky, ale třeba anglicky. A když tomu softwaru tuhle databázi předložíme a řeknete mu, že ony čtyři různé nahrávky patří jednomu člověku, on se naučí, jak se jejich projevy od sebe liší. Teď se snažíme o to, aby to uměl co nejpřesněji.

Kromě rozpoznávání hlasu telefonujícího člověka řešíte ještě jednu velmi zajímavou aplikaci, a to přepis řeči do písemné formy. Není to žádná nová věc, ale jen pár vědeckých týmů ve světě dokáže přepisovat spontánní mluvenou angličtinu, češtinu nebo nizozemštinu tak rychle a především přesně jako vy. K čemu je ona aplikace dobrá?