

## **POZOR, BOMBA!**

*Vlasta Radová*

Ve sdělovacích prostředcích se čas od času objeví zpráva o anonymních telefonátech oznamujících uložení bomby v nějakém objektu. Kriminalistům se původce takových telefonátů daří najít nejrůznějšími metodami. Lze například zjistit telefonní číslo nebo místo, ze kterého se volalo, ale lze také analyzovat hlas volajícího. Jak je možné poznat původce anonymních telefonátů na základě jeho hlasu?

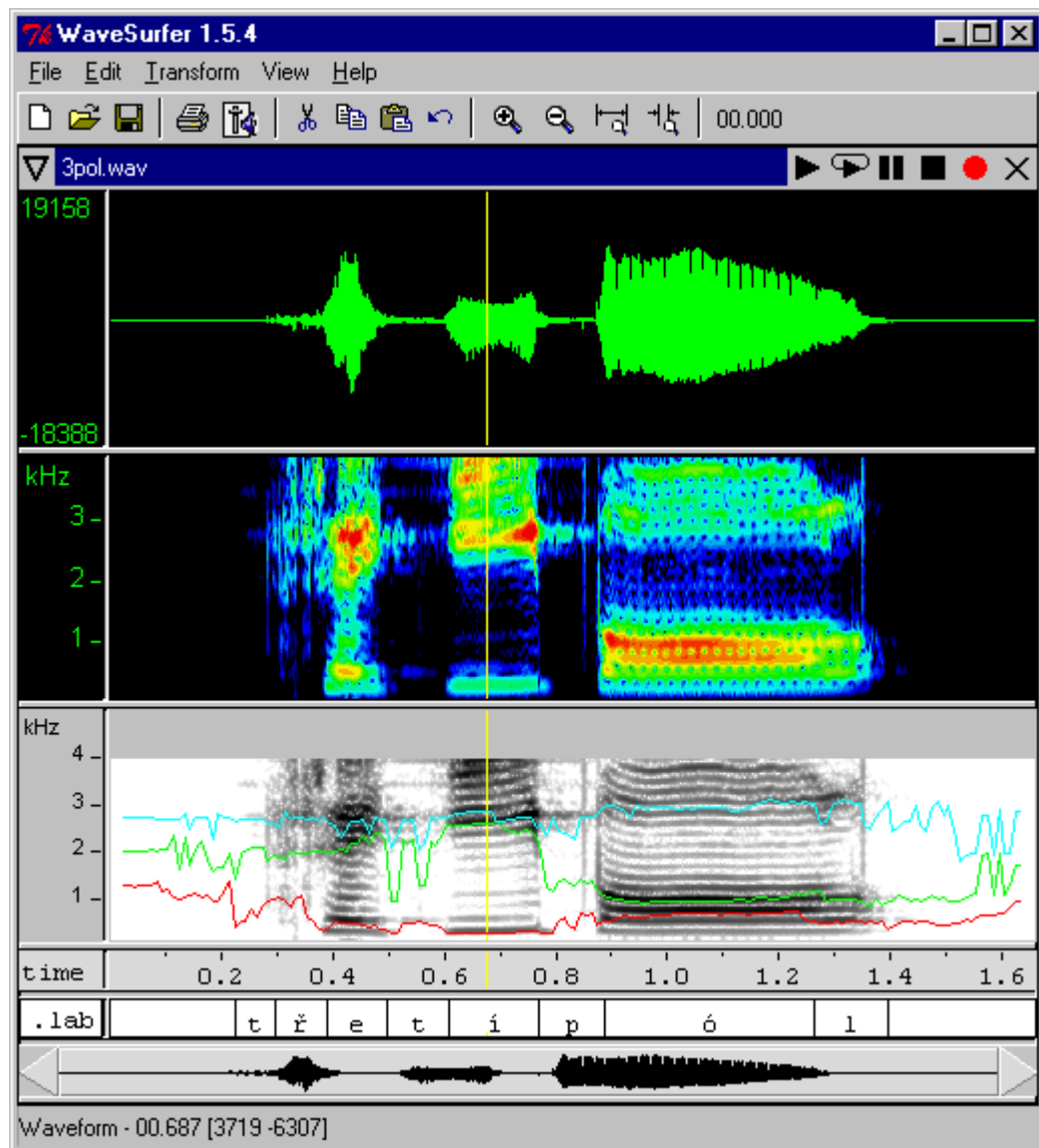
Fyzikální podstatou řeči je chvění vzduchu, které vnímáme jako zvuk. Zdrojem zvukové podoby jednotlivých hlásek je výdechový proud vzduchu z plic, který je modifikován artikulačními orgány člověka – hlasivkami, dutinou hrdelní, ústní a nosní, měkkým a tvrdým patrem, zuby, jazykem, čelistmi a rty. Hlasivky se pod tlakem výdechového proudu vzduchu při vytváření znělých hlásek rozkmitávají, čímž vzniká tzv. základní tón lidského hlasu. Kromě toho se v akustickém spektru hlásek objevuje řada vyšších zesílených tónů, které vznikají rezonancí v dutinách hlasového traktu. Těmto tónům se u samohlásek říká formanty, u souhlásek se označují jednoduše jako rezonanční kmitočty. Z hlediska vnímání řeči jsou pro určení samohlásek nejdůležitější první dva až tři formanty, které jsou pro každou samohlásku jiné. Vyšší formanty, které zůstávají pro všechny samohlásky stejné, obsahují značnou část informace o barvě hlasu řečníka.

Rozdíly v anatomické stavbě artikulačních orgánů jednotlivých lidí mají vliv na akustické vlastnosti řečového signálu a projevují se individuálními odchylkami hlavně v hodnotách frekvence základního hlasivkového tónu a ve frekvenci a šířce pásma jednotlivých formantů a rezonančních kmitočtů. Například frekvence základního hlasivkového tónu se u většiny lidí pohybuje při běžném hovoru v rozmezí 150 až 400 Hz, přičemž u žen je v průměru dvakrát vyšší než u mužů, a u dětí ve věku kolem deseti let se pohybuje okolo hodnoty 300 Hz. Podobně frekvence formantů jsou u žen asi o 17% a u dětí ve věku kolem deseti let asi o 25% vyšší než u mužů.

Spektrogram znázorňující spektrum řečového signálu daného člověka je tedy možné považovat za jakýsi „otisk hlasu“ tohoto člověka. Právě název „otisk hlasu“, který v analogii k pojmu „otisky prstů“ pro spektrogram poprvé použil Lawrence Kersta v roce 1962, značnou měrou přispěl k velmi rychlému rozvoji metod pro identifikaci člověka na základě hlasu. A to i přesto, že identifikace člověka na základě vizuálního porovnání dvou spektrogramů není příliš spolehlivá. Jedná se totiž o subjektivní metodu, která je silně závislá na zkušenostech odborníka, který spektrogramy porovnává. V následujících letech se tedy pozornost odborníků soustředila na vývoj spolehlivějších metod.

Současné metody rozpoznávání řečníka lze rozdělit do dvou skupin podle toho, jakou informaci pro rozpoznávání využívají. V řeči lze totiž najít v podstatě dva druhy informace, které charakterizují řečníka – charakteristiky nižší úrovně a charakteristiky vyšší úrovně. Charakteristiky nižší úrovně jsou takové charakteristiky, které lze změřit a číselně vyjádřit. Patří sem tedy například výše zmíněné hodnoty základního hlasivkového tónu a frekvence a šířky pásma formantů a rezonančních kmitočtů. Kromě toho sem patří také tempo řeči, které se projevuje v různé délce konkrétních slov i v časovém kolísání uvnitř slov, intonace a přízvuk, které lze popsat průběhem základního hlasivkového tónu podél určité promluvy, anebo hlasitost, která souvisí s amplitudou řečového signálu. K charakteristikám vyšší úrovně patří dialekt, styl řeči, a také specifické způsoby slovního vyjadřování jako třeba užívání konkrétních slov nebo konkrétních druhů smíchu.

Zatímco charakteristiky nižší úrovně lze poměrně snadno využít v automatických systémech rozpoznávání řečníka, charakteristiky vyšší úrovně se v takových systémech v současné době nevyužívají, protože je poměrně obtížné je číselně vyjádřit. Přesto mají při rozpoznávání člověka podle hlasu velký význam. Podle nich lze totiž usuzovat, jaké má daná osoba vzdělání a povolání, odkud pochází, a podobně. Využíváním těchto charakteristik v kriminalistice se zabývá věda, která se nazývá forenzní lingvistika, a experti z této oblasti dokáží významným způsobem zúžit okruh lidí, kteří přicházejí v úvahu jako možní pachatelé. Nevýhodou charakteristik vyšší úrovně však je, že jsou závislé na jazyku, jakým daná osoba hovoří. Proto se řada pracovišť po celém světě pokouší vyvinout metody, které by byly jazykově nezávislé. Tyto metody jsou založeny na pravděpodobnostním modelu řečníka, který se vytváří z charakteristik nižší úrovně získaných z dostupných promluv daného řečníka. V České republice se vývojem takového systému zabývá například katedra kybernetiky Západočeské univerzity v Plzni.



Obrázek znázorňuje různé reprezentace promluvy „třetí pól“. V horní části obrázku je zeleně zachycen přímo řečový signál, pod ním je uveden odpovídající spektrogram (červeně jsou znázorněny frekvence s největší amplitudou, modře frekvence s nejnižší amplitudou). Pod barevným spektrogramem je tentýž spektrogram znázorněn černobíle a barevně jsou v něm zvýrazněny průběhy prvních tří formantů, respektive rezonančních frekvencí. Pod časovou osou jsou pak pro informaci přibližně naznačeny hranice jednotlivých hlásek v promluvě.