

Systém Visualization Space Marka Lucenteho má nalézt své uplatnění například při vědeckých vizualizacích. Výkonný hardware poskytuje dostatečnou výpočetní sílu pro manipulaci s enormně velkými databázemi, například s geoseismickými daty. Mezi další oblasti využití patří vzdělávání a zábava. Uživatel se může zúčastnit virtuální procházky po historickém městě nebo letecké cesty nad kontinenty.

A jaké jsou ceny za Visualization Space? Systém sestává z PC, video digitizéru, zvukové karty, mikrofону a grafického akcelerátoru, to vše za cenu kolem 15 000 USD. Cena displeje záleží na jeho velikosti, zobrazovač o šířce 2 metry stojí kolem 30 000 USD.

Info: <http://researchweb.watson.ibm.com/imaging/vizspace.html>



# Rozhraní budoucnosti

Všechno bude jinak, rozhlašují počítačová vizionáři. Zahodte myši a klávesnice, přicházejí nová rozhraní mezi člověkem a strojem – s počítači budeme komunikovat prostřednictvím řeči, gestikulace, mimiky. *Michal Černý, michal.cerny@vogelburda.cz*

## Svět v jeho rukách

Vypadá to téměř jako zázrak. Experimentátor stojí v místnosti a na projekčním plátnu před ním se ve 3D stereoskopické projekci vznáší obraz zeměkoule. Muž na sobě nemá navlečené datové rukavice à la Tom Cruise ani nepotřebuje jakékoli trekovací zařízení – a přesto se zdá, jako kdyby „držel svět ve svých rukách“. Když pohne pravou paží, vyobrazený glóbus sleduje pohyb a přesune se z jednoho okraje projekčního plátna na druhý. „A teď se roztoč!“ přikazuje experimentátor a sleduje, jak glóbus pomalu začíná rotovat kolem své osy.

Sledujeme demonstraci systému Visualization Space od firmy IBM, který má ukázat novou cestu k budoucím rozhraním mezi člověkem a počítačem. Jeho autor, výzkumník Mark Lucente, o sobě prohlašuje, že nemá příliš rád typické proprietaty světa IT současnosti – klávesnice nebo monitory. Říká o nich, že jsou jenom malými průzory do digitálního prostoru. Lucente tak navrhl zcela nový způsob komunikace mezi člověkem a strojem – Visualization Space, interaktivní místnost, kde stačí mávnout rukou nebo vyslovit přání přirozenou řečí, počítač mluvené slovo nebo gestikulaci v mžiku zanalyzuje a v reálném čase zareaguje.

Mark Lucente navrhl svůj systém jako prostor o rozměrech přibližně 4 x 8 metrů, kdy jednu ze stěn tvoří plátno se zadní

projekcí. Uživatel se ocitá v interakci prostřednictvím dvou kanálů – jednak řeči, jednak přes vizuální vstup, v němž počítač analyzuje gestikulaci či postavení těla. Experimentátor má u sebe bezdrátový mikrofon, kterým se zaznamenává řeč, a také ho snímá kamera přenášející živé video. Pro analýzu gest a postojů použil tým výzkumníků software PersonFinder, jehož autory jsou vědci z MIT (Massachusetts Institute of Technology). Pro rozpoznávání řeči byl nasazen software IBM ViaVoice. →

## Zahraj country – a kde je bankomat?

IBM nabízí svoje systémy rozpoznání řeči automobilkám General Motors a Honda. V automobilech firmy Honda (např. v modelu Acura RL) systém dovoluje řidiči v průběhu jízdy zjišťovat pouze hlasem stav automobilu, kontrolovat rádio, CD přehrávač nebo řídit systém GPS, kontrolovat provoz a podobně. Jednoduchými řečovými příkazy může například vybírat oblíbené skladby na CD, volit a přepínat rozhlasové stanice, měnit hlasitost, ovládat mobilní telefon atd. To je velmi důležité zejména z hlediska bezpečnosti řízení, protože řidič nemusí sklánět svůj zrak a může se plně věnovat řízení automobilu a sledovat provoz před sebou.



## Virtuální asistent posádky

Smysluplná konverzace s roboty či palubním počítačem, která byla desetiletí jedním ze základních prvků literatury science fiction, se stává skutečností díky spolupráci vědců z Amesova výzkumného centra NASA a odborníků firmy Xerox. Cílem systému, který nese označení Clarissa, je usnadnit astronautům jejich každodenní pracovní úkoly na palubě International Space Station (ISS).

Astronauti během svého pobytu na mezinárodní vesmírné stanici vykonávají přibližně 12 000 úkonů, které zahrnují například údržbu systémů pro podporu života, kontrolu skafandrů, vědecké experimenty nebo pravidelné zdravotní prohlídky. Clarissa přitom reaguje na hlasové povely členů posádky a nahlas předčítá instrukce potřebné pro bezchybné plnění připravených úkolů.

Clarissa umí rozpoznat jednotlivá slova, ale také kompletní věty a slova v kontextu a dokáže reagovat na velké množství příkazů vyjádřených různým způsobem. Celý systém rozpoznávání řeči nyní zpracovává jednotlivá slova ve větě, definuje pravděpodobnost, podle které jsou jednotlivá slova správně identifikována, a následně aplikuje algoritmus se schopností učit se na základě již získaných informací.



### → Od nože z pazourku se máme co učit

Lze tedy v systému Visualization Space vidět budoucnost rozhraní mezi člověkem a strojem? Mnozí jsou přesvědčeni o tom, že ano – a že přesun ke skutečně „lidskému“ způsobu ovládání je už načase. „Navzdory technologickému pokroku, který dělí kamennou sekeru nebo nůž z pazourku od moderního počítače, se máme od těchto prastarých nástrojů stále co učit,“ míní v článku publikovaném na internetu novinářka Emily Benedeková. „Nástroje z minulosti byly mnohem těsněji adaptovány na své uživatele a jejich efektivnost závisela na tvaru a na zručnosti či síle těch, kterým sloužily. Počítače se naproti tomu vyvíjely poněkud odlišně. Původně byly vytvořeny pro vysoce speciální použití velmi poučených uživatelů, a teprve nedávno začaly být adaptovány pro běžný život.“

Mnozí proto tvrdí, že s technologiemi se budeme již brzy domlouvat gesty, řečí nebo obličejovou mimikou, a rozhlašují, že klávesnice u počítačů zmizí během několika málo desetiletí. Jedná se o realistickou předpověď? Vedoucí oddělení Voice Technologies and Systems firmy IBM Jan Šedivý říká, že je třeba být poněkud opatrnější. „Zcela určitě lze dnes říct, že interface budoucnosti bude multimodální – to znamená, že vstupy budou do systému proudit prostřednictvím různých modalit. Těmi mohou být například klávesnice nebo myš, může to být tužka píšící na obrazovku, která je „touch sensitive“ – nebo to může být i řeč.“

### Vize všudypřítomných senzorů

Firma IBM ve svých vizích vykresluje podobu budoucnosti, která by se v některých svých sociálních aspektech mohla zdát až neskutečná. Trendem by se mohlo stát rozpuštění počítačů do všudypřítomných senzorických polí, obklopujících lidské návštěvníky. „V současnosti digitální svět navštěvujeme, nahlížíme do něj světlíky počítačových monitorů a manipulujeme jím pomocí klávesnice a myši,“ píše k tomu William J. Mitchell ve své knize E-topia. „Počínáme si přitom podobně jako technik pracující s radioaktivním materiálem, jímž manipuluje v boxu pomocí návleků na ruce a umělých ramen. Naše stroje ovládají digitální svět přímo, ale málokdy si uvědomují analogový svět, který obklopuje jejich kyberprostor.“

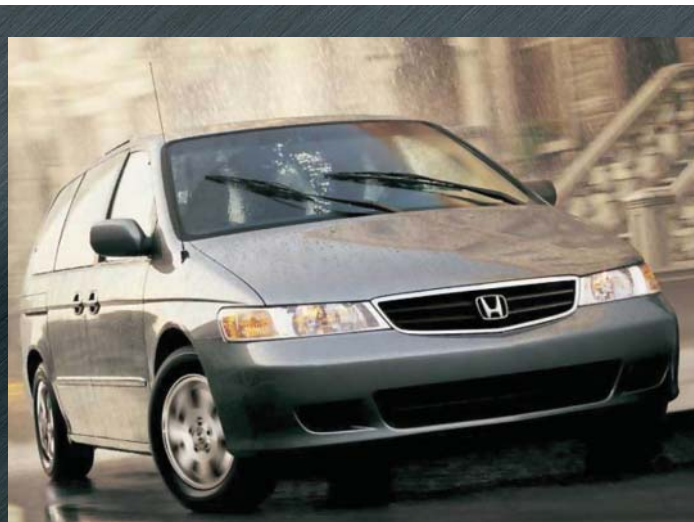
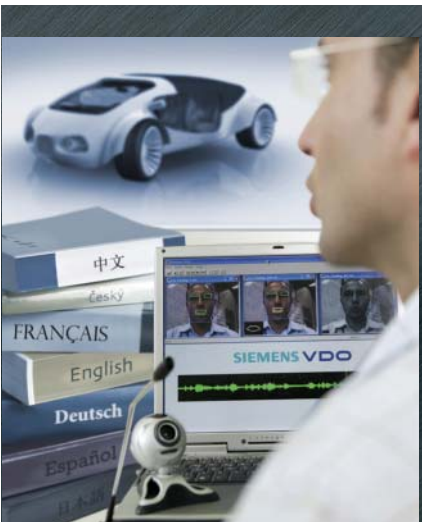
Podle IBM by se tak trendem mohly stát všudypřítomné technologie pronikající prostorem ve formě malých senzorů, které by

všude v „chytrých místech“ (smart places) sloužily uživatelům. Vezměme si například konferenci konající se v takovémto inteligentním prostoru. Videokamera by sledovala každého nově přichozího, systém by rozpoznal, o koho jde, a když by se dotyčný usadil na místo, projektor by na stůl před ním mohl promítnout třeba program, který se bude na schůzi projednávat.

„Samozřejmě se nabízí otázka, jestli by lidé skutečně chtěli být sledováni senzory a zda takové vymoženosti skutečně chceme mít, nebo ne,“ dodává Jan Šedivý. „Každá technologie bezpochyby naráží také na sociálně-psychologické bariéry.“

### Zaostřeno na řeč

Jednou z modalit, která by mohla sehrát významnou úlohu při utváření vyspělého interface budoucnosti, je přirozená řeč. Jak již →





→ bylo zmíněno, v systému Visualization Space Marka Lucenteho tuto úlohu obstarává program Via Voice, jeden z nejrozšířenějších softwarů pro tento segment trhu. Takzvané NLU neboli Natural Language Understanding je záležitostí teprve několika posledních desetiletí, boom řečových systémů nastal přibližně na konci 80. let minulého století, v souvislosti s uvedením do praxe statistických metod (skryté Markovovy modely), s nimiž software pro rozpoznávání řeči pracuje.

Zřejmě každého, kdo by se zajímal o technologie rozpoznávání řeči, napadne otázka: Co všechno již počítače dokáží rozpoznat? Jedná se pouze o relativně omezené sady příkazů, nebo jsou skutečně schopny porozumět plynulému toku řeči? I s použitím moderních matematických metod z oblasti umělé inteligence musí dialog se strojem stále pojednávat o poměrně úzce ohraniče-



„Interface budoucnosti bude multimodální.“

Jan Šedivý, ředitel Voice Technologies and Systems IBM

ném tématu. Jádrem NLU aplikace nejsou ani tak rutiny určené pro samotné rozpoznávání řeči, jako spíše tzv. dialog manažer – softwarová nadstavba, která uživateli pokládá upřesňující dotazy, systém dále filtruje komunikaci od redundantních šumů a vyjádření a z odpovědi je schopen generovat informace pro databázi.

### Jak se dostanu k nejbližší benzinové pumpě?

Systémy NLU, schopné porozumět přirozené řeči a vybavené dialog manažery, nacházejí své uplatnění již dnes, a to především v dvou hlavních oblastech. IBM nabízí své produkty ViaVoice pro telefonii v call centrech. Počítačové programy dnes úspěšně nahrazují například spojovatelky. U leteckých nebo dopravních společností lze díky NLU systémům objednávat lety nebo poskytovat informace o spojení v jízdních řádech.

A druhá oblast využití? Především se jedná o mobilní zařízení – ať už mobilní telefony, či hlasové systémy v automobilech. V případě mobilních telefonů se hlas zatím příliš neprosadil, v automobilovém průmyslu je však situace zcela jiná. Řidiči se musí plně věnovat řízení a každé rozptýlení jejich pozornosti znamená snížení bezpečnosti – v tomto případě tedy ovládání řeči, bez nut-

Jednou z oblastí, kam firmy směřují svoje řečové produkty, je telefonie v call centrech. Počítačové systémy, které jsou vybaveny softwarovými nadstavbami – dialog manažery – dokáží již dnes úspěšně nahrazovat lidské operátory. To pro firmy znamená značnou úsporu nákladů.



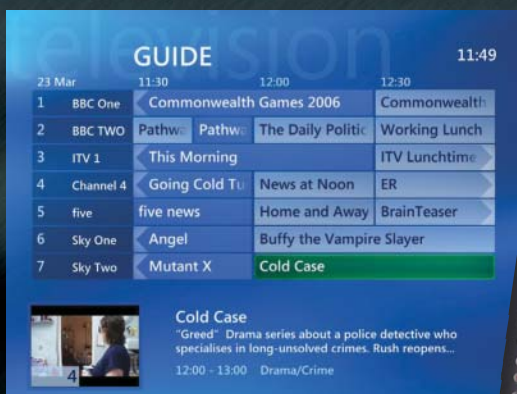
nosti snímat ruce z volantu, představuje značný přínos. Dokazují to automobilky, které stále častěji vybavují své modely navigačním GPS systémem ovládaným hlasem.

### Odfiltrovat šum stále ještě moc neumíme

Vyplní se tedy vize, že během několika málo let přejdeme na interface ve stylu Visualization Space Marka Lucenteho a zahodíme myši a klávesnice? V jakýchkoli futurologických predikcích je třeba být opatrný a mít na paměti, že při použití multimodálních systémů je každý typ vstupu velmi užitečný ve specifických situacích, zatímco v jiných může být téměř nepoužitelný. „Například řeč je velmi vhodná, pokud chceme pracovat s velkými agregátními celky,“ říká Jan Šedivý. Klávesnice umožňuje velmi efektivně zadávat čísla nebo jiné údaje. Pokud však váš televizor přijímá 400 televizních stanic se 2000 show, je mnohem efektivnější jednoduše říct: „A teď mi ukaž program, kde hrají detektivku s Jackem Nicholsonem.“

Samotné strojové zpracování řečových vstupů dnes nepředstavuje závažnější problém, hardware je dostatečně výkonný. Potíž však často vyvstává s mikrofony – pokud se uživatel pohybuje například v běžném domácím prostředí, které vyplňuje hluk ostatních členů rodiny, předpokládá se, že nebude potřeba, aby u sebe měl navíc vlastní mikrofon. Systém zabudovaný ve zdech domu by měl splnit jakékoli dostatečně nahlas vyřčené přání. Z toho důvodu existují mikrofonová pole, pomocí nichž je možné uživatele trekovat. I tak ovšem stále existuje problém odfiltrování nežádoucích šumů, s nimiž si současné systémy často nedokáží poradit. „Dá se říci, že na tom intenzivně pracujeme,“ říká Jan Šedivý. I z toho důvodu, že umělá inteligence našich systémů není zcela dořešenou záležitostí, patří skutečně dokonalý interface budoucnosti z velké části do říše vědecké fikce.

Michal Černý ■



**CO BUDE V TELEVIZI?:** V souvislosti s příchodem digitální televize bude problém vyznat se v nabídkách desítek stanic. Cílem je udělat ovládání hlasem, aby bylo možné třeba se jednoduše zeptat: na kterém programu teď běží zprávy?



**TECHNOLOGIE OD FIRMY OMRON:** Počítače či mobilní telefony budeme stále častěji ovládat gesty či obličejovou mimikou. Předzvěstí v tomto směru je technologie rozpoznávání obličej. OKAO Vision Face Recognition Sensor má identifikovat tvář uživatele mobilního telefonu a zajistit tak vyšší bezpečnost mobilních zařízení.