

11

Počítačové zpracování zvuku

VÝUKOVÝ TEXT PRO STUDENTY INFORMAČNÍCH TECHNOLOGIÍ

Marek Lučný, SŠPU OPAVA 2015

Zvuk je mechanické vlnění v látkovém prostředí, které je schopno vyvolat v lidském uchu sluchový vjem. Frekvence tohoto vlnění leží v rozsahu přibližně 16 Hz až 20 kHz; za jeho hranicemi člověk zvuk sluchem nevnímá. V širším smyslu lze za zvuk označovat i vlnění s frekvencemi mimo tento rozsah. Například zvuk s frekvencí nižší než 16 Hz (který slyší např. sloni) nazýváme **infrazvuk**. A zvuk s frekvencí vyšší než 20 kHz (např. delfini vnímají zvuk až do frekvencí okolo 150 kHz) nazýváme **ultrazvuk**.

Akustika je věda, která se zabývá zvukem a jeho šířením. Zvukové vlny se mohou šířit ve hmotném prostředí, nejčastěji bývá vodičem zvuku vzduch. Zvuky se šíří i kapalinami (např. vodou) nebo pevnými látkami (např. stěnami domu); ve vakuu zvuk zachytit nemůžeme.

Zdrojem zvuku může být každé kmitající těleso. Pravidelné kmity tvoří **tóny** (hudební zvuky), nepravidelné kmity označujeme jako **hluky** (nehudební zvuky). Původcem hudebních zvuků mohou být lidské hlasivky nebo různé hudební nástroje, hluky mohou způsobit např. vítr, výstřel, běžící motor apod.

Základní charakteristikou zvuku je jeho **frekvence** – počet kmitů za sekundu (jednotkou je 1 hertz - Hz). Jak už bylo uvedeno, většina lidí vnímá frekvence v rozsahu 16 Hz až 20 kHz. Nízké frekvence vnímají lidé jako hluboké tóny, vyšší frekvence jako vysoké tóny; s rostoucím věkem klesá vnímavost zejména vyšších tónů. Nejcitlivěji vnímá lidské

ucho frekvence v rozsahu 2 – 4 kHz, který je nejdůležitější pro srozumitelnost řeči. Nejvyšší informační hodnota řeči je přenášena v pásmu 0,5 – 2 kHz.

Každý zvuk, hudební i nehudební, se vyznačuje také fyzikální intenzitou, s níž je rovnocenná veličina nazývaná **hladina intenzity zvuku** (měřená v dB). Dynamický rozsah, rozdíl mezi nejhlasitějším a nejtichším vnímatelným zvukem, je uprostřed slyšitelného frekvenčního pásma asi kolem 120 dB. Na okrajích pásma je však mnohem menší. Hranici velmi tichých zvuků (přibližně kolem 20 dB) nazýváme **prahem slyšitelnosti**, zvuky silnější než 120 dB (velmi hlasitá hudba, startující letadlo, výbuch...) jsou za tzv. **prahem bolesti** a mohou způsobit i trvalé poškození sluchu.

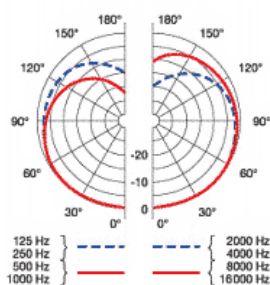
Vedle těchto dvou základních vlastností zvuku rozlišujeme také **barvu zvuku**. Díky ní jsme od sebe schopni rozlišit zvuk vydávaný různými hudebními nástroji, i když právě hrají stejný tón. Každý hudební zvuk je totiž možné rozložit na tzv. **harmonické složky**; se základním, nejsilnějším tónem, souznějí doprovodné slabší zvuky jiných frekvencí. Čím více takových harmonických složek daný tón obsahuje a čím výrazněji tyto složky znějí, tím je barva tónu plnější, sytější. Barvu tónu u jednotlivých hudebních nástrojů můžeme ovlivňovat tvarem a materiálem rezonanční skříňky. Tím se zesilují nebo zeslabují určité vyšší harmonické tóny.

Akustický zvukový signál můžeme zachytit pomocí mikrofону. **Mikrofon** je zařízení, které snímá zvuk a převádí akustické vlny na **elektrický analogový zvukový signál**.

Kondenzátorový mikrofon patří k nejkvalitnějším snímačům zvuku a používá se v nahrávacích studiích. Funguje na principu změn elektrického napětí, které způsobuje měnící se kapacita kondenzátoru. Mikrofon je tvořen membránou z kovu či pokovené plastové fólie. V případě **elektrodynamického mikrofónu** pohybuje membrána cívkou v magnetickém poli, čímž je vytvářen elektrický proud. Dynamické mikrofóny jsou méně citlivé než kondenzátorové, lépe proto zpracují například hlasitý zpěv při živých vystoupeních, ozvučení veřejných shromáždění apod. Bývají poměrně odolné proti mechanickému poškození a nevyžadují napájení. Problémem je jistá frekvenční závislost – při snímání zvuku z větší vzdálenosti se mění jeho frekvence a dochází ke zkreslení.

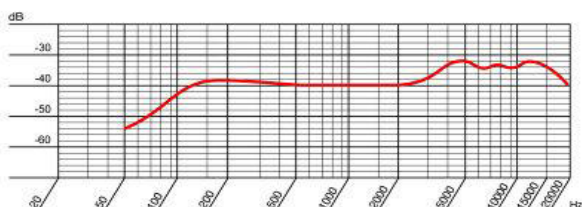
K důležitým parametrům mikrofónu patří:

Směrová charakteristika. Ukazuje, s jakou citlivostí mikrofon snímá zvuky přicházející v ose a mimo osu mikrofónu. Toto se zaznamenává do tzv. kruhového diagramu; bývá tak často zaznamenáno více směrových charakteristik (většinou jinou barvou) daného mikrofónu pro různé frekvence. Rozeznáváme mikrofóny se směrovou charakteristikou kulovou (všesměrovou), osmičkovou, kardioidní (ledvinová)...



Citlivost mikrofónu je poměr výstupního napětí mikrofónu a akustického tlaku, který toto napětí vybudil; udává se v milivoltech na Pascal (mV/Pa).

Frekvenční charakteristika znázorňuje citlivost mikrofónu pro zvuk přicházející zepředu v ose mikrofónu pro různé frekvence (nejčastěji slyšitelné spektrum 20 Hz – 20 kHz).



Digitalizace zvuku je proces, při kterém dochází k převodu analogového zvukového signálu (spojitý) na digitální (číslicový, nespojitý). Teprve digitalizovaný zvuk můžeme zpracovávat v počítači a uchovávat na digitálních médiích (CD, DVD, pevné disky, flash paměti, paměťové karty). Převod signálu z analogového na digitální (např. při vstupu z mikrofónu, zesilovače), nebo naopak z digitálního na analogový (např. při výstupu do reprobeden, sluchátek) zajišťuje **zvuková karta**, jejíž nedílnou součástí je speciální čip obsahující **převodníky A/D a D/A**.

Základem digitalizace zvuku je tzv. **vzorkování (sampling)**. Spočívá v rozdělení souvislého analogového signálu na sérii diskrétních (nesouvislých, časově oddělených) stavů. To znamená, že ve stanoveném časovém intervalu je zjištěn a zaznamenán aktuální stav signálu (vzorek). Čím kratší je tento interval, tím vyšší je vzorkovací frekvence a tím je digitální záznam bližší původnímu analogově zachycenému signálu. Platí přitom pravidlo (Shannon-Kotělnikovův teorém), že vzorkovací frekvence (**SF – sampling frequency**) by měla být minimálně dvakrát vyšší než nejvyšší frekvence původního signálu. Chceme-li tedy zaznamenat jakýkoliv zvuk v oblasti slyšitelné pro lidské ucho (16 Hz až 20 kHz), musí být vzorkovací kmitočet minimálně 40 kHz. Například u standardního zvukového CD je vzorkovací kmitočet 44,1 kHz; u čistě zvukových formátů, jako je například DVD-Audio, se využívá vzorkovací kmitočet 48 kHz, případně jeho násobky (96 kHz, 192 kHz pro formáty s vysokým rozlišením).

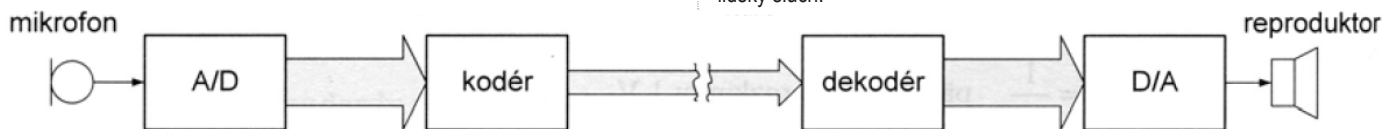
Druhou částí procesu digitalizace je **kvantování**, určení rozsahu přenášeného signálu v dynamické rovině. Pro každý vzorek je zaznamenána síla zvuku v binární podobě (tzv. **bitová hloubka převodu**). Pro tradiční audio CD je standardní bitová hloubka 16 bitů ($2^{16} = 65\,536$ úrovní), v telefonech se využívá 8bitové rozlišení, naopak v novějším formátu DVD-Audio může být každý vzorek uložen ve 24 nebo 32 bitech. Protože na každý bit připadá 6 dB dynamického rozsahu, maximální teoretická dynamika pro 16bitový signál je 96 dB, pro 24bitový signál pak 144 dB.

Negativním důsledkem digitalizace může být tzv. **kvantizační zkreslení**. Pro jednotlivý vzorek zvuku totiž musí převodník zvolit konkrétní hodnotu z nabízené škály úrovní (např. v případě 16bitové hloubky jednu ze 65 536 hodnot) a pochopitelně to nemusí být přesně ta hodnota, která by odpovídala původnímu analogovému signálu. Mnohými úpravami digitálního zvuku se bohužel kvantizační zkreslení zhoršuje.

Digitální zvuková data jsou ve svém plném rozsahu velmi objemná.

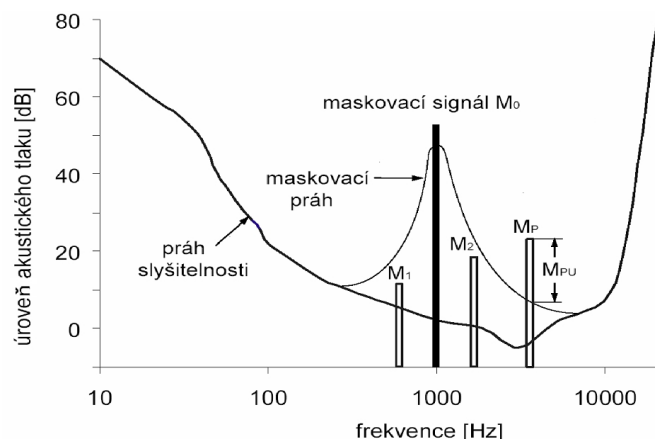
Například pouhých 10 sekund zvukového záznamu v kvalitě CD zabere v nekomprimované podobě téměř 2 MB paměti ($44\,100$ (vzorkovací frekvence) * 2 (16bitová hloubka) * 2 (levý a pravý kanál) * 10 (sekund) = $1\,764\,000$ B).

Výrazného snížení objemu zvukových dat dosáhneme použitím **komprese** - potlačením dat, která jsou *redundantní* (nadbytečná) a *irelevantní* (nevýznamná). Komprese může být **beztrátová** (po dekompresi získáme původní data), nebo **ztrátová** (dochází při ní k ztrátě některých dat).



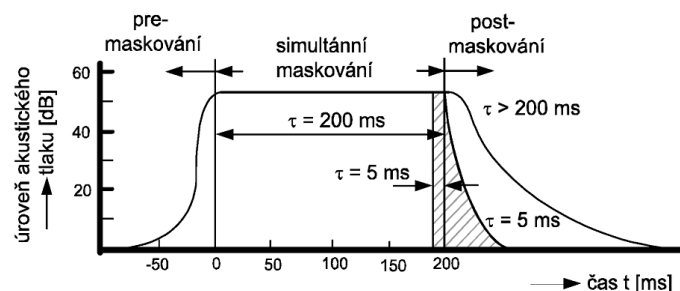
Na základě psychoakustického modelu lze pro ztrátovou zvukovou kompresi použít tzv. **maskování signálu**.

Frekvenční maskování. Je založeno na faktu, že člověk běžně není schopen rozlišit dva frekvenčně blízké tóny. Pokud znějí dva tóny současně, může jeden z nich potlačit slyšitelnost toho druhého. Silnější zvuk (na jakékoliv frekvenci) vytvoří tzv. maskovací práh. Zvukové vzorky, které nepřesáhnou úroveň maskovacího prahu, proto nemusí být přenášeny - ucho člověka je totiž stejně nezaznamená. Příkladem může být zvuk sirény, který svou intenzitou spolehlivě překryje všechny do té doby slyšitelné zvuky blízké frekvence.



Ztrátové kompresní algoritmy využívají tzv. **psychoakustického modelu**. Ten vychází z vědeckých poznatků o omezených schopnostech lidského vnímání zvuku. Lidské ucho totiž není schopno zachytit všechny složky reprodukováného zvuku (například ze dvou po sobě jdoucích tónů s podobnou barvou zvuku slyšíme dobře jen ten hlasitější), a navíc přijímá pouze akustické signály určitého frekvenčního pásma (např. zvuky nad 16 kHz jsou pro člověka s přibývajícím věkem slyšitelné jen slabě, nebo vůbec). Rovněž vnímání intenzity zvuku je závislé na jeho frekvenci - člověk dokáže nejcitlivěji reagovat na frekvence mezi $1\text{--}5\text{ kHz}$. Psychoakustický model tyto lidské slabiny zná, matematicky je definuje a shrnuje výsledky do tabulek. Při komprimaci porovnávají kodéry výstupní signál s tabulkou a odebírají ty nejméně významné informace pro lidský sluch.

Časové maskování. Silný zvuk neomezuje vnímání „slabších“ zvuků jen v kmitočtové oblasti, ale i v časové. Pokud zazní silný signál, je naše vnímání dalších zvuků omezeno na několik desítek až stovek milisekund i po jeho ukončení a slabší signály opět neslyšíme. Doba dozívajícího maskovacího efektu je závislá na době, po kterou signál zněl a pokud zněl krátce (např. 5 ms), je kratší i dozívající maskovací efekt. Tato metoda se nazývá *post-maskování*. Ovšem díky komplikovanému chování lidského mozku lze použít i tzv. *pre-maskování*, kdy je na několik milisekund zamaskován dokonce i slabší signál, který předcházal zaznění intenzivnějšího zvuku.



Kromě maskování signálu využívají ztrátové kompresní algoritmy i další pokročilé matematické postupy, které umožňují efektivní úsporu toku dat. Účinnost algoritmů se však silně odvíjí i od struktury dat, které hudební soubor reprezentují. Zjednodušeně se dá říci, že pokud zvolíme pro kompresi hudby stejný kodek i stejný datový tok, bude se velikost výsledného souboru významně lišit i podle stylu hudby. S jiným žánrem se zásadně změní zastoupení frekvenčních složek a jejich relativní důležitost pro kvalitní reprodukci.

Zvuková data ukládáme do souborů v určitých **formátech**. Na začátku audiosouborů je **hlavička** obsahující údaje o délce záznamu, o rychlosti přehrávání (resp. *sample rate*), o počtu bitů potřebných k uložení jednoho vzorku, o počtu kanálů (*mono*, *stereo*). Součástí hlavičky mohou být i podrobnější textové informace o zvukovém záznamu, tzv. **ID3 tagy** (např. název interpreta hudební skladby, hudební žánr, rok vzniku apod.).

V případě komprimovaných formátů musí být uveden použitý **kodek** (převzato z anglického *codec*, složenina začátků slov „kodér“ a „dekodér“). Jde o program, který na základě jistého algoritmu zajišťuje kompresi (část programu nazývaná *encoder*) a dekompresi (*decoder*) dat. Existuje celá řada audio kodeků; k beztrátovým patří např. **FLAC** nebo **Monkey Audio**, k ztrátovým řadíme třeba **libvorbis** nebo **WMA**.

Mezi nejpoužívanější kodeky patří **LAME**, jenž je vyvíjen jako open source a používán jako enkódér pro formát MP3. Umožňuje nastavení komprimace vícekanálového zvuku:

- Volba **stereo** zajistí samostatné kódování levého i pravého kanálu.
- Volba **joint stereo** využívá faktu, že na obou kanálech bývá obvykle

stejný, nebo velmi podobný signál. Speciální algoritmus proto věrně zaznamená jeden kanál, zatímco v druhém kanálu jsou zachyceny pouze informace o rozdílu s druhým kanálem. V tomto módu dojde ke zvýšení celkové kvality na úkor částečné ztráty sterea.

Jiné důležité nastavení je spojeno s **datovým tokem** (*bitrate*, určen počtem kilobitů za sekundu), který rozhoduje o výsledné kvalitě komprimovaného signálu. Platí, že čím je menší datový tok, tím je menší soubor a nižší kvalita zvuku. V případě MP3 se nejvíce používá *bitrate* v rozmezí $128\text{ až }320\text{ kbps}$. Od 160 kbps je záznam téměř nerozeznatelný od originálu.

Konstantní datový tok (**CBR** - *constant bitrate*) se v čase nemění a nepřípustně se charakteru zvukové nahrávky. Méně zatěžuje procesor, ale obvykle znamená o něco větší soubor.

Proměnný datový tok (**VBR** - *variable bitrate*) je závislý na charakteru zvuku - v klidných místech je nižší, zvyšuje se u intenzivnějších partií. Řešení je náročnější na výpočetní výkon, ale je šetrnější k paměti a zvukový záznam bývá kvalitnější.

Průměrný datový tok (**ABR** - *average bitrate*) je vypočítán jako suma přenosové rychlosti každého vzorku dělena počtem vzorků. Skladba má v průměru nastavený jeden *bitrate*, jsou však možné výkyvy.

NEKOMPRIMOVANÉ FORMÁTY

WAV (*Waveform audio file format*). Základní zvukový formát určený původně pro OS Windows, ale dnes podporovaný většinou programů pro práci se zvukem. Lze v něm zaznamenat digitální zvukovou stopu v jakékoliv kvalitě (vzorkovací frekvence, bitová hloubka, mono či stereo). Umožňuje jednoduchou manipulaci a editaci zvuku. Na druhou stranu jsou z tohoto důvodu WAV soubory poměrně objemné, nepraktické pro ukládání i pro sdílení na Internetu - 1 minuta v CD kvalitě (16bit / 44kHz) zabírá přibližně 10 MB.

AIFF (*Audio Interchange File Format*). Starší, ale stále ještě podporovaný formát zvuku ve zcela nekomprimované podobě původně používaný v počítačích Apple a Silicon Graphics. Stále je podporován současnými přehrávači iPod.

FORMÁTY S BEZZTRÁTOVOU KOMPRESÍ

FLAC (*Free Lossless Audio Codec*). Open source zvukový kodek, který pro konverzi zvuku na nižší datový obsah využívá lineární predikci a samotný princip by se dal přirovnat k fungování univerzálních komprimačních programů jako WinRAR či WinZip. Velikost komprimovaného zvukového souboru je rovna zhruba 50 – 60 % originálu.

Monkey's Audio. Velikost komprimovaných souborů je oproti originálu zhruba poloviční, ale oproti konkurenci je o něco pomalejší. Soubory tohoto formátu mají příponu *ape* a jsou podporovány i některými osobními přehrávači.

FORMÁTY SE ZTRÁTOVOU KOMPRESÍ

MP3. Nejrozšířenější zvukový formát ztrátové komprese, který se zrodil ve Fraunhoferově institutu. Je založen na kompresním algoritmu MPEG-1 s kó-

dovacím schématem layer 3, z něhož vychází i samotný název. Zdokonalený formát **MP3Pro** přinesl především technologii **SBR** (*Spectral Band Replication*) pro zlepšení kvality zvuku a odstranění nežádoucích artefaktů při použití úspornějších datových toků. V podstatě v sobě nese dva proudy: první obsahuje standardní MP3 a druhý kóduje vyšší frekvence.

OGG Vorbis. Open source ztrátový audio kodek, který měl ambice nahradit právě MP3. Oproti MP3 využívá lepší akustické modely i matematické principy a pro stejnou úroveň kvality zvuku potřebuje zhruba poloviční datový tok. Neschází mu možnost vícekanálového zvuku, ale nepodporuje synchronizaci ani ochrany proti chybám.

AAC (*Advanced Audio Coding*). Ztrátové kompresní schéma multikanálového zvuku, které je navrženo jako nástupce MP3 a standardizované v MPEG-2 i MPEG-4. Na vývoji se podílel Fraunhoferův institut, Dolby Labs, Bell Labs, Sony a Nokia. Nejedná se o jednotný formát a existuje celá řada modifikovaných enkodérů lišících se zejména zvukovou kvalitou. Používán a podporován je mimo jiné celou řadou zařízení od Apple (iPhone, iPod), Sony (PlayStation), některými telefony Nokia, herní konzolou Wii atd. AACplus představuje rozšíření standardního AAC o velmi pokročilé kompresní technologie včetně SBR v oblasti vysokých kmitočtů (10 -20 kHz) pro ještě lepší kvalitu zvuku při nízkých datových tocích.

WMA (*Windows Media Audio*). Vyspělý kompresní formát pro zvuková data z dílny Microsoftu. Postupem času bylo vytvořeno více různých variant úzce specializovaných kodeků například pro kompresi hlasu a elektronických knih (**WMA Voice**), archivaci v bezeztrátové podobě (**WMA Lossless**) či kódování vícekanálového zvuku (**WMA Pro**).

Zvuk může být na počítači vygenerován rovněž uměle; hovoříme v takovém případě o **zvukové syntéze**. Možností, jak elektronickou cestou generovat vlnový průběh, který lze následně přeměnit na zvuk, je několik; nejčastěji se setkáme s FM syntézou a wavetable syntézou.

FM syntéza vychází z úvahy, že hudební zvuky mají podobu cyklu, který se skládá ze čtyř částí: nástupu, poklesu, trvání a doznívání zvuku (*model ADSR = Attac-Decay-Sustain-Release*). Podstatou syntézy je určit vlnový průběh daného hudebního nástroje zadáním hodnot pro každou z částí cyklu. V tomto případě jsou do zvukového formátu zaznamenávány informace o nástroji, který tón hraje (piano, housle, varhany ...), o výšce tónu, délce tónu, dynamice úhozu na klávesu, případně další charakteristiky.

Wavetable syntéza využívá zvuky skutečných nástrojů, jež jsou uloženy v paměti ROM (*Read Only Memory*) zvukové karty, ze které jsou následně přehrávány podle stisknuté klávesy. Každý zvuk a jeho výška má své číslo a jsou organizovány do jakési tabulky. Proto název *WaveTable* - tabulka vlnových průběhů. Touto syntézou lze s vysokou věrností docílit akustických zvuků.

Obvykle je FM syntéza realizována obvodem **MIDI** (*Musical Instrument Digital Interface* = digitální rozhraní pro hudební nástroje). MIDI je mezinárodní standard používaný v hudebním průmyslu jako elektronický komunikační protokol, který dovoluje hudebním nástrojům, počítačům i dalším přístrojům komunikovat v reálném čase prostřednictvím definovaného sériového rozhraní.

Pro zpracování MIDI-dat je nabízena celá řada aplikací a pro jejich uložení je používán formát **SMF** (*Standard MIDI File*). Tyto soubory mají obvykle příponu *.mid*. Jedná se o formát vektorový, který stejně jako v případě grafiky, umožňuje snadné opravy a úpravy vytvořených souborů bez ztráty kvality. Zvukový soubor MIDI je zaznamenán v podstatě tak, jako by byl zapsán do notové partitury. Jednotlivým zvukům je stanovena výška, délka a hlasitost tónu, tempo skladby a odpovídající zvuk nástroje. Z polyfonie těchto zvuků (většinou 16 nebo 32 hlasů najednou) potom vzniká výsledná skladba.

Výhodou tohoto formátu je takřka zanedbatelná datová velikost (i rozsáhlá skladba může mít jen několik kilobitů). V tomto formátu není možné zaznamenávat lidskou řeč nebo jiné přirozené zvuky. Uvedený způsob se poměrně snadno kóduje a umožňuje vytvářet velmi malé soubory. Zjednodušující povaha modelu ADSR však může ubírat hudební věrnosti, samozřejmě také v závislosti na kvalitě používané zvukové karty.

SOFTWARE PRO ZÁZNAM A EDITACI ZVUKU

Tyto aplikace se zaměřují na záznam a následnou editaci zvukového signálu. Umožňují zvuková data nahrávat, přehrávat, importovat i exportovat v různých formátech. Nabízejí základní operace potřebné k editaci zvukových souborů včetně mixování stop a různých efektů. Mohou obsahovat i nástroje sloužící k analýze zvuku (např. konfigurovatelný spektrogram, analýzu frekvence) nebo k filtraci nežádoucích zvuků (odstranění šumu, praskání na záznamech ze starších desek apod.).

Aplikace: *SoundForge*, *Cubase*, *CoolEdit*, *Adobe Audition*, *Audacity* atd.

SOFTWARE PRO PRÁCI S MIDI

Slouží k aranžování hudby za použití tónů jednotlivých nástrojů. Nahrávku v MIDI formátu můžeme pořídit buď použitím speciálního hardware (klávesy připojené přes MIDI rozhraní k počítači), nebo využitím tzv. **notátoru** – programu, který nabízí notový zápis partů zvolených nástrojů.

Aplikace: *ProTools*, *GuitarPro*, *Rosegarden*, *Anvil Studio* a další.

SOFTWARE PRO TRACKOVANOU HUDBU

Používá se pro tvorbu počítačové hudby postavené na řetězení krátkých zvukových vzorků (samplerů) do několika stop.

Aplikace: *Renoise*, *Reason*, *FL Studio*, *Linux Multimedia Studio (LMMS)*...

PŘEHŘÁVAČE AUDIO SOUBORŮ

Kromě přehrávání zvukových souborů (včetně streamovaných médií) nabízejí mnohé přehrávače také řadu dalších funkcí – např. sestavování playlistů, prohlížení a editaci ID3 tagů, online informace (např. zobrazování obalů desek), rippování (uložení zvukových stop z CD v podobě datových souborů) atd.

Aplikace: *Windows Media Player*, *WinAmp*, *Audacious* a jiné.

SOFTWARE PRO KONVERZI AUDIA

Slouží pro převedení zvukových záznamů do jiných formátů.

Aplikace: *Format Factory*, *Media Coder*, *XrecodeII* apod.

Streaming (z anglického *stream* – proud) je technologie kontinuálního přenosu audiovizuálního materiálu mezi zdrojem a koncovým uživatelem. Na rozdíl od stahování (*downloadu*) se video nebo audio soubor při streamingu pouze přehrává a neukládá do počítače. Přenos dat může probíhat v přímém přenosu (on-line), nebo ze záznamu (off-line). Ke streamingu audia se využívá různých formátů i kodeků (WMA, MP3, OGG, AAC+) zpravidla v datových tocích od 16 do 256 kbps. Audio může být mezi serverem a klientem streamováno jako **single bitrate**, čili jako jeden konstantní datový tok, nebo **multibitrate**, kdy stream obsahuje více konstantních datových toků. Program přehrávající multibitrate stream ze serveru dokáže automaticky měnit kvalitu zvuku v případě zhoršení (nebo naopak zlepšení) kvality internetového připojení posluchače.

Podcasting je metoda šíření informací do jisté míry konkurující rádiu. Slovo podcasting vzniklo spojením názvu přehrávače iPod (*Pod* = *Personal On Demand*) a anglickým slovem *broadcasting* (vysílání). Jde o zvukové nebo video záznamy, jež autor podcastu umísťuje na Internet v podobě souborů (často ve formátu MP3) a na které odkazuje na webových stránkách, především však v uzpůsobeném RSS feedu. Ten pak specializovaný program (zvaný *podcatcher* nebo *podcast receiver*) průběžně monitoruje a nové soubory sám stahuje a nahrává do uživatelského osobního přehrávače. Mezi významné přednosti podcastingu patří snadné poskytování obsahu jednoduchým umístěním zvukového souboru na Internetu a možnost uživatele přehrát si soubor kdykoliv, nezávisle na programovém schématu rádia.

Rozpoznávání řeči (*speech recognition*) se rozumí automatický převod mluvené řeči do textu. Metody rozpoznávání řeči jsou citlivé na rozdíly v intonaci i výslovnosti jednotlivých mluvčích, což má být překonáno pomocí složitých výpočetních operací. Proces rozpoznání řeči (zvláště plynulé řeči) je proto velmi náročný na výpočetní výkon.

Rozpoznávání řeči může být **závislé na mluvčím** (*SD* - *Speaker Dependent*), nebo **na mluvčím nezávislé** (*SI* - *Speaker Independent*). Systémy, které jsou závislé na mluvčím, dosahují lepších výsledků pro danou osobu. Je to ovšem podmíněno tím, že mluvčí musí pořídit několik hodin nahrávek vlastního hlasu, na základě nichž jsou poté vytvářeny modely pro rozpoznávání. Systémy nezávislé na mluvčím vycházejí z nahrávek hlasů velkého množství lidí ve snaze o vytvoření obecnějšího modelu.

Programy využívající počítačové rozpoznávání řeči mohou být využity pro vytváření textových dokumentů, pro přepis televizních a rozhlasových pořadů i pro hlasové ovládání počítače, které je výhodné zejména pro handicapované osoby. K nejpokročilejším komerčním programům pro angličtinu (a další jazyky) patří *Dragon Naturally Speaking*, pro češtinu např. *NovaVoice* či *Newton-Dictate*. Rozpoznávání řeči je v češtině (a obecně ve všech ohebných jazycích) podstatně složitější hlavně kvůli mnohem rozsáhlejšímu slovníku, který musí obsahovat všechny možné varianty slov (výsledky skloňování, časování). Úspěšnost programů na rozpoznávání řeči se liší podle druhu textu, ale zpravidla přesahuje 90 procent.

Syntéza řeči (*Text-to-Speech Synthesis*) je proces, jehož cílem je co nejvěrněji reprodukovat lidskou řeč konkrétního člověka, a to i výškou, zabarvením a tónem mluvy. Využívá se zde technologie **TTS** (*text-to-speech*), která s pomocí speciálních algoritmů nejprve text analyzuje a normalizuje, poté dochází k transformaci do podoby výslovnosti. Při převodu textu na mluvenou řeč se za sebe kladou podle kontextu drobné fonetické prvky, z nichž se skládá plynulá řeč. Předčítaný text se analyzuje z hlediska melodiky a rytmu řeči (*prosodie*) a podle textu se modifikuje také zvukový charakter hlasu. Výsledkem procesu je syntetická řeč velmi blízká řeči přirozené.

Syntetizovaná řeč může nahradit skutečného lidského řečníka v širokém spektru oblastí – od rutinního oznamování opakujících se informací (zastávky MHD, nádraží, apod.), přes hlasový monitoring údajů (řidičí střediska), informační a dialogové systémy (automatická spojovatelka, telefonní informační linky), až po vysoce propracované a přirozené čtení libovolných textů (e-mail, SMS, celé knihy). Stále častěji se syntéza řeči uplatňuje v zábavním průmyslu.

Nejužitečnějšími jsou nepochybně aplikace pro handicapované, které mohou využívat lidé s různým zrakovým, sluchovým nebo hlasovým postižením. Pro nevidomé je určena čtečka obrazovky (screen reader), hlasově postižení mohou k vytváření „své“ řeči využívat svůj „osobní“ systém syntézy řeči. Lidé, kteří ztratili řeč po mozkové mrtvici, mohou pro výuku řeči využít technologii založenou na modelu mluvčího hlavy. Sluchově postižení mohou upřednostnit poslech srozumitelné a akusticky čisté syntetické řeči před dynamickou přirozenou řečí.

MIKROFON

Volba mikrofonu. Je vhodné upřednostnit mikrofon kardioidní (směrová charakteristika ve tvaru srdce) před kulovým či osmičkovým. Studiový kondenzátorový mikrofon má lepší kmitočtovou charakteristiku oproti levnějšímu dynamickému mikrofonu.

Umístění mikrofonu. V případě dynamického mikrofonu se doporučuje jeho umístění co nejblíže ke zdroji zvuku. Je-li to možné, nahrávejte zvuk externím mikrofonem s dostatečně dlouhým kabelem, v opačném případě je třeba co nejvíce utlumit hluk samotného nahrávacího zařízení (např. dekou).

Ochrana proti větru. Naráží-li na membránu mikrofonu silnější proud vzduchu (vítr, ale i proud vzduchu při vyslovování retnic - hlásky p, t, b, ...) vznikají nepřijemné hluboké pazvuky znehodnocující záznam. Speciální „rozštěpený“ obal (tzv. „mrtvá kočka“) zajistí rozptýlení soustředěného proudu zvuku i za cenu zhoršeného snímání vysokých frekvencí.

NAHRÁVACÍ MÍSTNOST

Odhučnění místnosti. Nedokonale utěsněné dveře a okna mohou způsobovat pronikání hluku z vnějšku. Seběmenší skvira může propouštět hluk, proto je nezbytná izolace, nejlépe gumová.

Tvar místnosti. Problém je především s rohy v místnosti a s rovnoběžnými stěnami. Vhodné je zbavit se pravých úhlů, např. pomocí paravánů udělat

z místnosti víceúhelník. Nevhodný může být i nízký plochý strop, který způsobuje třepetavou ozvěnu.

Akustika místnosti. Důležitý je vyrovnaný dozvuk v celém frekvenčním pásmu; zvuk má být všude dostatečně silný a podle možnosti alespoň přibližně stejně silný, jednotlivé zvuky lidské řeči a krátce trvající hudební tóny nesmějí splývat. Vzdálenost od nejbližší odrazné plochy (stěny) by měla být 3–8 metrů. Pokud to není možné, je třeba zajistit obložení stěn: pro ztlumení vysokých frekvencí používáme hrubě omítnutou zeď, tapety, koberce, závěsy, záclony, kartony od vajíček; pro hluboké kmitočty větší nábytek, sádkartonové přepážky, dřevěné desky s odstupem od zdi apod.

TECHNICKÁ OPATŘENÍ PŘI NAHRÁVÁNÍ

Připojení periférií. Připojením jakékoli periférie se na kostru počítače může dostat rušivé napětí a v důsledku toho pak i brum do nahrávky. Do zvukové karty zapojíme jen ty konektory, které jsou v dané chvíli potřeba. Zabráníme tím vzniku zpětné vazby mezi navzájem propojenými audio přístroji.

Nastavení zvukové karty. Na softwarovém ovladači zvukové karty vypneme (funkce MUTE) vstupy, které momentálně nepotřebujeme. Pokud se do zvukové karty dostává hluk z CD mechaniky, tak ji při záznamu zvuku na PC nepoužíváme.