

Ročníkový projekt – specifikace

Syntéza zvuku ze spektrogramu

Jan Krajíček
Vedoucí: Martin Bálek

1. června 2009

Cílem tohoto projektu je napsání programu, který by uměl vytvořit spektrogram zvukové nahrávky dle zadaných parametrů a se znalostí těchto parametrů ze spektrogramu zpětně vytvořit zvukovou nahrávku, podobnou té zdrojové. Jde o ztrátový proces a výsledek bude oproti původní nahrávce zkreslený, ale měl by být stále dobře rozeznatelný. Program bude mít jednoduše použitelné grafické rozhraní a poběží na platformách Linux a Windows.

1 Popis funkcionality

1.1 Fourierova transformace

Významným nástrojem pro práci se zvukovými signály je Fourierova transformace. Spektrogram s lineární osou frekvencí je možné získat téměř bezprostřední aplikací na signál „po částech“, což je tzv. STFT (Short Time Fourier Transform). V podstatě jde o rozklad úseků signálu do frekvenčních složek, z vodorovného vektoru určitého počtu vzorků získáme svislý vektor komplexních čísel, jejichž absolutní hodnota určuje intenzity harmonických signálů o různých frekvencích. Provedení absolutní hodnoty samozřejmě znamená ztrátu informace. Šířka jednotlivých frekvenčních pásem (v Hz) je v tomto případě konstantní. Rozlišení frekvenční osy (tedy počet, resp. velikost pásem) je dána právě počtem vzorků. Čím vyšší rozlišení chceme dosáhnout, tím delší časový úsek je nutno vzít. Zde dochází k další ztrátě informace, protože celý zvolený úsek se promítne do jediného svislého pruhu spektrogramu.

Další informací, která je při transformaci ztracena, je fázový posun získaných harmonických signálů – zachycena je jen jejich intenzita. Pro reprodukci zvuku nicméně není fázový posun důležitý.

1.2 Tvorba spektrogramu ze zvuku

Spektrogram zvukové nahrávky je třírozměrný graf zobrazující intenzitu frekvenčních složky zvuku v čase. Čas se obvykle zobrazuje na vodorovné ose, frekvence na svislé a intenzita je značena zbarvením bodu.

Program bude umět tento graf vytvořit dle parametrů, které nastaví uživatel.

Pro účely zachycení zvuku podobně tak, jak ho slyší člověk (a tedy pro lepší výsledek zpětné syntézy), je vhodné osy frekvencí i intenzity kreslit logaritmické. Ve spektrogramu s lineární osou frekvencí jsou nízké tóny vykresleny do velmi malého prostoru, přestože jsou důležitou složkou slyšitelného zvuku. Naopak méně důležité vysoké tóny dostávají zbytečně velký prostor. Při použití lineární osy intenzity je pak špatně zachycen velký rozsah hlasitostí (výsledný spektrogram je příliš tmavý).

Důležitým parametrem spektrogramu je rozlišení frekvenční a časové osy, které na sobě přímo závisí – vyšší rozlišení jednoho znamená nižší rozlišení druhého. Tato volba je nevyhnutelným kompromisem a pro různé zvuky jsou vhodné různé hodnoty, volba vhodné hodnoty je na uživateli.

Program bude pro uložení spektrogramu nabízet volbu z běžně používaných formátů obrázků, zejména png a bmp (jpeg také, ale ten není vhodnou volbou z důvodu ztráty kvality).

Pro spektrogram bude možno zvolit libovolnou paletu barev pro znázornění intenzity, bude tedy možno vytvářet černobílé i barevné spektrogramy.

1.2.1 Tvorba lineárního spektrogramu

Spektrogram s lineární osou frekvencí je tvořen pomocí Short Time Fourier Transform, která je popsána výše. Při tvorbě lineárního spektrogramu je podstatná volba následujících parametrů:

- Velikost okna: Okno je počet vzorků, ze kterého vznikne jeden svislý pruh spektrogramu. Jak je popsáno výše, použitím většího okna lze získat jemnější rozlišení na frekvenční ose (rozsah frekvencí v jednom vodorovném pruhu), ale zhorší se rozlišení na časové ose (tedy délka zvuku, který bude zobrazen v jednom svislém pruhu). Pro různé účely jsou vhodné různé hodnoty.
- Míra překrytí oken: Jednotlivá okna vzorků se obvykle překrývají – jeden vzorek se promítne do více svislých pruhů. Větší překrytí dává méně zrnitý, ale širší spektrogram.
- Okenní funkce: Pro ještě přirozenější zobrazení se na překrývajících okna může použít okenní funkce, která ztlumí signál na okrajích okna (kde se překrývá se sousedním), ale nezmění ho (nebo jen málo) uprostřed. Často se používá např. Gaussova funkce. Jeden vzorek se tedy promítne do více pruhů v různých intenzitách.

Spektrogram se potom získá aplikací Fourierovy transformace na jednotlivá okna – výsledné hodnoty jsou přímo vykresleny do spektrogramu.

1.2.2 Tvorba logaritmického spektrogramu

Logaritmický spektrogram bude tvořen podobně, jako to dělají analogové přístroje. Místo rozdělení na krátké časové úseky, jako je tomu u lineárního spektrogramu, se rozdělí frekvenční osa na několik částí a každá je zachycena vlastním filtrem, přičemž šířka každého filtru (rozsah zachycovaných frekvencí v Hz) je dvojnásobkem šířky předchozího. Tyto filtry jsou postupně aplikovány na zvukový signál a amplitudy výsledných částí signálu tvoří vodorovné pruhy spektrogramu.

Obdobou velikosti okna v lineárním spektrogramu je zde šířka jednotlivých filtrů, měřeno v oktávách na pruh. Tyto filtry se také obvykle překrývají, a jejich tvar může být upraven nějakou okenní funkcí.

1.3 Syntéza zvuku ze spektrogramu

Program bude umět z výše popsaného spektrogramu vytvořit zpetně zvuk. Spektrogram může pocházet z tohoto programu, nebo z jiných zdrojů (např. i

ručně nakreslený), pokud jsou známy jeho parametry.

Z důvodu ztrátovosti převodu zvuku do spektrogramu i zpětného procesu bude výsledná nahrávka nevyhnutelně zkreslená, nicméně by měla být pořád dostatečně podobná např. na to, aby šlo rozumět takto zpracovanému mluvenému slovu, nebo identifikovat hudební skladbu.

Výslednou zvukovou nahrávku bude možno uložit minimálně do formátů ogg (s kompresí) a wav (bez komprese).

1.3.1 Popis syntézy

Největší ztrátu informace při převodu do spektrogramu způsobuje spojení mnoha vzorků zvuku do jednoho sloupce, který představuje obvykle desítky milisekund. V těchto intervalech je snadné vzorky zrekonstruovat, ale pro vytvoření poslouchatelného zvuku je potřeba o několik řádů více vzorků. Běžná nahrávka jich má 44,1 za milisekundu. Chybějící vzorky je třeba co nejlépe odhadnout.

Z tohoto důvodu jsou pro syntézu vhodnější spektrogramy s menším počtem vzorků na sloupec (a tedy větším počtem sloupců) – ztráta rozlišení na frekvenční ose (do jisté míry) neovlivní výslednou kvalitu tak výrazně, jako menší počet vzorků, ze kterých se dá při syntéze vyjít. Bude vhodné zjistit optimální parametry spektrogramu pro dobrou reprodukci hudby nebo hlasu a tyto hodnoty nabídnout uživateli jako výchozí.

Přímočarý způsob jak získat chybějící vzorky mezi dvěma sloupci spektrogramu je interpolace z harmonických signálů, které okolní sloupce popisují. V blízkosti intervalů známých vzorků bude reprodukce poměrně přesná, ale ve větší vzdálenosti mezi vzorky může být zvuk velmi zkreslený. Tato metoda syntézy by měla podávat uspokojivé výsledky a přitom být poměrně rychlá. Najít vhodný způsob interpolace bude pro dobrý výsledek nezbytné.

Jiný způsob syntézy, který bude program nabízet, je použití náhodného šumu jako výchozího zvuku, oproti harmonickým signálům jako v předchozím případě. Tento šum je rozložen sadou frekvenčních filtrů, podobně jako signál při převodu do spektrogramu. Výsledky programu ARSS naznačují, že nejvhodnější je růžový šum (narozdíl od neutrálního bílého šumu jsou v něm hlasitější nízké, basové frekvence), nicméně bude typ šumu pravděpodobně ponechán jako volba. Tato metoda syntézy může vytvořit přirozenější zvuk než syntéza pomocí harmonických signálů a pro obecné použití bude pravděpodobně vhodnější, i když se zašumění nevyhnutelně do výsledku promítne. Parametry, které se ukáží jako nejlepší budou opět nabídnuty jako výchozí.

1.4 Uživatelské rozhraní

Program bude mít grafické rozhraní ve kterém bude přímo možné načíst zvuk a převést ho dle zvolených parametrů do spektrogramu, který bude následně zobrazen a bude ho možno uložit.

Podobně půjde načíst obrázek – spektrogram, zvolenou metodou pak bude převeden do zvuku, který půjde přímo přehrát nebo uložit.

Pro usnadnění práce s parametry spektrogramu bude, pokud to výstupní formát obrázku podporuje, při převodu zvuku do spektrogramu možno uložit jeho parametry jako komentář v metainformacích obrázku. Program si je pak pro zpětnou syntézu načte, takže je nebude muset vyplňovat uživatel.

2 Existující software

Z existujícího softwaru umí syntézu spektrogramu provádět zejména program ARSS¹ s kódem dostupným pod licencí GNU GPL. Tento program má konzolové rozhraní, pracuje pouze s formáty wav a bmp, které jsou pro svou velikost poměrně nepraktické, a není již vyvíjen, protože po verzi 0.2.3 přešel v komerční Photosounder² s poněkud širším záběrem.

Tento program by měl implementovat funkcionalitu ARSS s příjemným grafickým rozhraním a prací s běžně používanými formáty, aby s ním mohl experimentovat i problematiky neznalý člověk. Navíc bude možné pracovat i se spektrogramy s lineární frekvenční osou.

3 Technické prostředky

Program bude napsán v jazyce C++, pro snadnou kompilaci na různých platformách bude použit systém CMake³. Cílové platformy jsou Windows a Linux.

Knihovny, které budou použity:

- Qt⁴ pro grafické rozhraní
- FFTW⁵ pro rychlou fourierovu transformaci
- libsndfile⁶ pro práci se zvukem

¹Webová stránka ARSS: <http://arss.sourceforge.net>

²Webová stránka Photosounderu: <http://photosounder.com>

³Webová stránka CMake: <http://www.cmake.org>

⁴Webová stránka Qt: <http://www.qtsoftware.com/products>

⁵Webová stránka FFTW: <http://www.fftw.org/>

⁶Webová stránka libsndfile: <http://www.mega-nerd.com/libsndfile/>