

VZDÁLENÁ ANALÝZA LIDSKÉHO HLASU - BEZEZTRÁTOVÉ NAHRÁVÁNÍ ZVUKU PŘES IP SÍŤ.

Tomáš Kulhánek, Marek Friš, Milan Šárek

Anotace

Aplikace ParVRP vyvinutá Výzkumným centrem hudební akustiky HAMU v prostředí Matlab umožňuje vytvoření fonetogramu a další následné analýzy jakými jsou plocha hlasového pole, poloha nejčtenějších zastoupení, statistiky hlasových projevů. V příspěvku představíme nasazení aplikace ParVRP ve virtualizovaném systému XEN a přístupnou přes protokol RDP v síti CESNET 2 a obecně Internetu. Během vývoje jsme vyzkoušeli několik možností pořizování nahrávek ze zvukové karty klientského počítače ke vzdálené aplikaci, neboť tato možnost není standardně v protokolu RDP k dispozici. RDP protokol verze 7.0 dostupný od nejnovějších systémů Windows Server 2008 R2 přenáší zvuk pomocí ztrátové komprese, která vede ke zkresleným výsledkům následné analýzy hlasu. Modifikace rozšířeného protokolu RDP v 5.2 pomocí Sound over RDP přináší podobný problém zkresleného zvuku přeneseného ke vzdálené aplikaci. Proto jsme vyvinuli bezzeztrátovou možnost nahrávání zvuku přes IP síť, která přináší stejnou kvalitu pořízeného záznamu jako na lokálním počítači, oproti předchozím ztrátovým kompresím je zhruba 10 krát náročnější na propustnost sítě a pro potřeby průběžné realtime analýzy potřebuje propustnost skoro 1 MBit/s, která je dosažitelná v běžných v lokálních sítích a je běžně dosažitelná i v organizacích připojených do sítě CESNET2

Klíčová slova:

fonetogram, vzdálená plocha, RDP, přenos zvuku

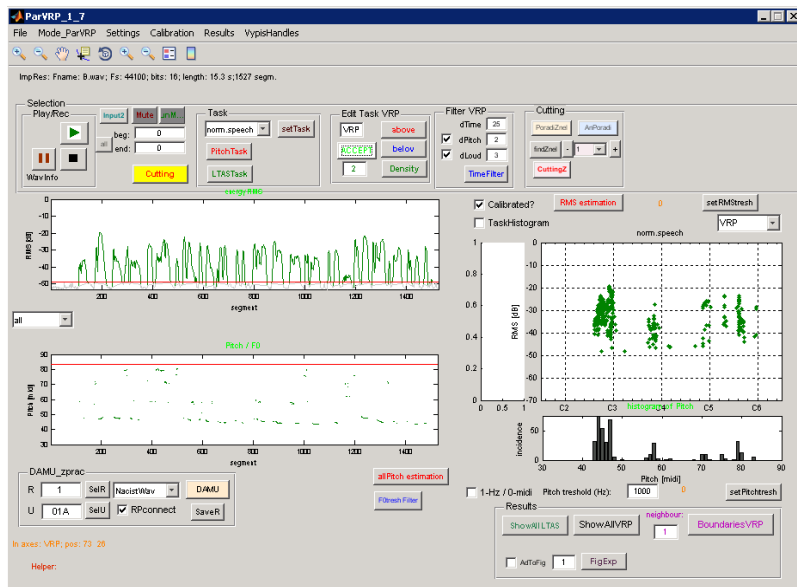
Úvod

Ve spolupráci Hudební akademie múzických umění (HAMU), Ústřední vojenské nemocnice, Pardubickou krajskou nemocnicí a sdružením CESNET z.s.p.o. pracujeme na projektu vzdálených vyšetření v oblasti otorhinolaryngologické medicíny[1], konkrétně v oblasti pořízení a analýza hlasového signálu. Ve výzkumném centru hudební akustiky HAMU jsme vyvinuli aplikaci ParVRP [2], která provede analýzu hlasového signálu z pořízených nahrávek, nebo za běhu aplikace s využitím mikrofonu připojeného ke zvukové kartě počítače.

Abyste tuto aplikaci mohli využívat spolupracovníci z různých míst republiky, umístili jsme aplikaci do virtuálního prostředí v rámci již existující pilotní infrastruktury, která byla budována v minulosti pro potřeby gridového projektu Globus MEDICUS a zpřístupnili ji pomocí protokolu RDP (Remote Desktop Protocol). Tato koncepce se obecně nazývá tenkým klientem. Klientský počítač má jednoduchý prohlížeč a veškerá aplikační logika je provozována na straně „tlustého“ serveru. Protokol RDP od verze 5.2 přenáší od klienta události z klávesnice a změny polohy myši k serveru, která tyto události vyhodnotí

a vykreslí je v prostředí, či předá aplikace a změny grafického vzhledu jsou zakódovány a přeneseny zpět ke klientovi, který je vykreslí na obrazovku uživatele. Protokol RDP verze 5.2 umí přeměňovat zvuk ze vzdálené aplikace ke klientovi na zvukovou kartu, ale neumí přeměňovat nahrávání zvuku z klientského mikrofону k serverové aplikaci. Tato funkcionality je dostupná až od RDP verze 7 dostupné od MS Windows 2008 R2, případně lze tuto funkcionality doinstalovat pluginem třetí strany. Nicméně v minulosti jsme ukázali, že obě dostupná řešení pro přeměňování zvuku z mikrofону k serveru pomocí protokolu RDP jsou ztrátová v kvalitě, dochází k ořezání vysokých či nízkých frekvencí – i v kvantitě, dochází ke ztrátám některých úseků nahrávky. Tudíž nejsou příliš vhodná pro přesnou analýzu lidského hlasu[3]. Navíc tyto a další metody např. VoIP technologie či obecně přenosu zvuku přes IP síť jsou navrhovány s požadavkem realtime přenosu a poslechu zvuku, tudíž odchylky v rychlosti přenosu zvuku vyrovnávají dynamicky dočasným zhoršením kvality tím i objemu přenášené zvukové frekvence. V našem případě je balík audio dat jen balík obyčejných dat, které chceme přenést beze ztrát kvality a jedno v jakém pořadí pakety k cíli dojdou a jestli a k jakému dojde zpoždění.

V této práci představíme řešení přenosu zvukového signálu z mikrofону lokálního počítače až k aplikaci, které rozšiřuje protokol RDP, používající k přenosu zvuku vyhrazený RDP kanál s možností použití nastavitelné komprese audio signálu a to hlavně bezetrátově.



Obrázek 1: Obrazovka aplikace parametrického fonetogramu ParVRP s analýzou hlasového pole v pravé části okna

Analýza zvuku

Ve Výzkumném centru hudební akustiky HAMU bylo vytvořeno uživatelské prostředí pro analýzu hlasových signálů ParVRP – parametrizované hlasové pole [3] pracující v postprocessingovém módu. Aplikace využívá numerické knihovny systému Matlab. Program umožňuje ze zvukových dat ve formátu WAV analyzovat základní frekvenci (pomocí autokorelační funkce), vypočte okamžitou energii signálu a spektrální charakteristiky pomocí rychlé Fourierovy transformace. Vypočtená data jsou následně efektivně filtrována podle vlastností lidského hlasu, tak aby vytvořila konečné zobrazení hlasového pole (voice range profile, VRP) a jeho parametrů. Systém hlavně umožňuje efektivní segmentaci signálů na jednotlivé typy vyšetřovaných hlasových úkonů.

Přesměrování nahrávání přes RDP

Na virtuálních serverech se systémy MS Windows 2003 a MS Windows 2008 provozujeme aplikaci ParVRP, která se automaticky spustí při přihlášení uživatele na vzdálenou plochu.

Protokol RDP jsme modifikovali na klientské straně pluginem, který čeká na signál k zahájení a ukončení nahrávání a souběžného přenosu zvuku přes vyhrazený virtuálního kanálu v navázaném RDP spojení a na serverové straně samostatným modulem otvírajícím virtuální kanál.

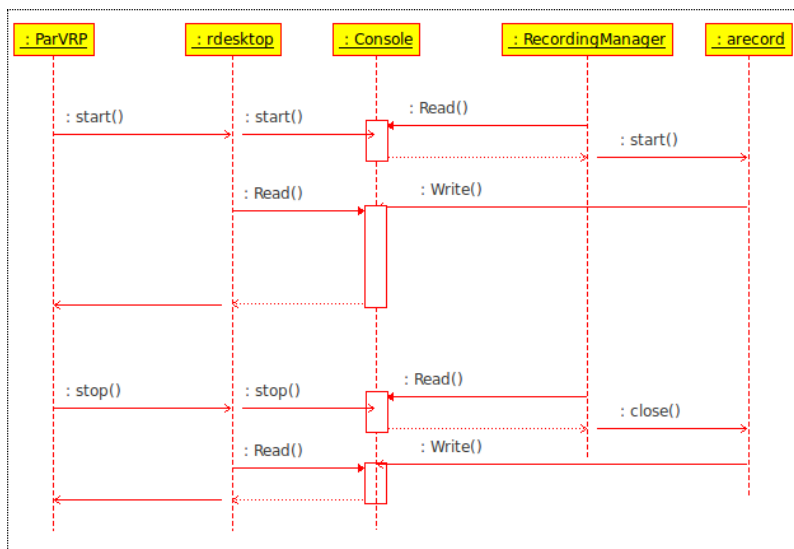
Na serverové straně běží modul, který posílá signál k zahájení a k ukončení nahrávání a přijímá binární data nahrávky standardně ve formátu WAV a buďto je ukládá do souboru, nebo je předává aplikaci ParVRP k dalšímu zpracování. Modul používá funkce z knihovny Remote Desktop Service API (Wtsapi32.dll) a otvírá pojmenovaný virtuální kanál, který je používán pro posílání signálu a příjem dat [6].

Plugin na klientské straně jsme vyvinuly pro platformu Windows i Linux. Plugin na klientské straně v Linuxu je realizován jako aplikace, která čte zprávy ze standardního vstupu a zapisuje binární data na standardní výstup. Modifikovali jsme RDP klienta rdesktop pomocí záplaty přidávající nový parametr:

```
-r addin:<channelname>:</path/to/executable>[:arg1[:arg2:]...]
```

Ten spustí program a veškerou komunikaci z RDP kanálu <channelname> přesměrovává na standardní vstup programu a výstup programu přesměrovává zpět do RDP kanálu.[5] Sekvenční diagram na *Obrázku 2.* ukazuje sekvenci při zahájení a ukončení nahrávání. Plugin jsme implementovali jako třídu RecordPlugin. Ze serveru přijde asynchronní zpráva „start“, přes standardní výstup (objekt Console) tato zpráva doputuje k třídě RecordPlugin, která vyvolá nahrávání s předvolenými parametry tak, že výsledný WAV je v průběhu nahrávání již přesměrováván na standardní výstup. Ten je kódován do RDP kanálu k serveru. Pokud server chce ukončit nahrávání, pošle signál stop, který se obdobným způsobem distribuuje až k třídě RecordPlugin. Zpátky se ještě pošle ukončovací sekvence WAV. Serverový modul i klientský plugin musí poslouchat tentýž kanál, který je identifikován svým jménem. Prototyp

klientského pluginu i serverového modul jsou napsány v C# a zkompileované do .NET. Klientská část v Linuxu využívá k vykonávání .NET aplikací projekt MONO [7] a pro digitalizaci nahrávání na klientovi arecord[8].



Obrázek 2: Sekvenční diagram při zahájení a ukončení nahrávání pomocí pluginu *rdesktop*.

Výsledky

Pokud je zapnuto nahrávání z klientské strany, datový tok z klienta k serveru dosahuje průměrně 98 kB/s. Pokud je klient připojen pomalejším připojením, dochází ke kumulaci neposlaných dat na klientské straně a ke zpoždění doručení dat, které ale nemá vliv na kvalitu pořízeného záznamu. Pro srovnání „sound recording redirection“ v protokolu RDP v 7 je datový tok od klienta k serveru průměrně 10 kB/s a obdobné výsledky dává „Sound over RDP“ v protokolu RDP v 5.2.

Závěr

Přenos zvuku pořízeného na lokálním počítači ke vzdálené aplikaci pomocí virtuálních kanálů v protokolu RDP je efektivním řešením pro systémy sledující koncept tenkého klienta i pro systémy, které chtějí obecně přenést balík dat z klienta na server. Toto řešení využívá již navázané RDP spojení, tudíž není nutné navazovat a konfigurovat nové extra síťové spojení pro přenos audio dat a konfigurovat firewally na cestě. Toto řešení však není vhodné např. pro využití v IP telefonii, neboť plýtvá kapacitou sítě a nereaguje na výkyvy

v latenci IP sítě. Uvedený způsob přenosu audio signálu je náročný k přenosové kapacitě sítě, oproti řešením se ztrátovou kompresí využívající různé kodeky je až 10x náročnější. S uvažovaným nasazením v rámci vysokorychlostní sítě CESNET2 jsou však tyto nároky zdaleka naplněny s dostatečnou rezervou. Díky bezeztrátovému přenosu audio signálu je tento typ nasazení aplikace na vzdálený server a zpřístupnění lokálních zdrojů včetně nahrávání z lokálního mikrofonu plnohodnotnou náhradou za lokální instalaci na klientském počítači. Toto řešení je navíc akceptovatelné pro zamýšlené nasazení této služby v klinické praxi, kdy pacient nemusí dojíždět na specializované pracoviště, případně specialisté nemusejí dojíždět na různá pracoviště, ale vyšetření hlasu může být prováděno na dálku.

Plánujeme systém rozšířit o plugin pro klientský prohlížeč vzdálené plochy na platformě MS Windows a doplnit možnost komprese při přenášení audio dat.

Poděkování

Tento příspěvek vznikl za podpory projektu Výzkumného centra hudební akustiky (za podpory Ministerstva školství a mládeže, ČR Pr.No. 1M6138498401) na Hudební fakultě Akademie múzických umění v Praze a s podporou výzkumného záměru MŠMT „Optická síť národního výzkumu a její nové aplikace“ ve sdružení CESNET z.s.p.o.

Literatura

- [1.] Šárek M., Kulhánek T., *Nové směry medicínských aplikací sdružení CESNET, MEDSOFT 2009.* (Milena Ziehamlová Ed.) Praha: Agentura Action M, Praha 2008, str. 145-148. ISBN 978-80-904326-0-4
- [2.] Frič M., *Parametrizovaný fonetogram obecných řečových a hlasových projevů – ParVRP, MARC-Technologický list č. 12, Zvukové studio HAMU, Praha 2007, <http://web.hamu.cz/zvuk/vyzkum/dokumenty/TL12x.pdf>*
- [3.] Kulhánek T., Frič M., Šárek M.: *Vzdálený přístup k virtuálním výukovým a výzkumným aplikacím - podpora foniatrických vyšetření, MEFANET 2009 sborník přednášek, Brno, ISBN: 978-80-7392-118-7*
- [4.] *Remote Desktop Protokol [http://msdn.microsoft.com/en-us/library/aa383015\(VS.85\).aspx](http://msdn.microsoft.com/en-us/library/aa383015(VS.85).aspx)*
- [5.] Yakimenko S., *“Writing plugins for Rdesktop”, The Code Project, Apriorit Inc, 1.12.2009, Web, <http://www.codeproject.com/KB/cross-platform/writing-plugins-for-rd.aspx>*
- [6.] Selvin, *How to Write a Terminal Services Add-in in Pure C#, The Code Project, 14.11.2006, Web, <http://www.codeproject.com/KB/system/TSAaddinCS.aspx>*
- [7.] Mono, *Cross platform, open source .NET development framework, Web, 2.2.2010, http://www.mono-project.com/Main_Page*
- [8.] Kysela J., *arecord, Linux manual page, 2.8.2001, Web, http://linuxcommand.org/man_pages/arecord1.html*

Kontakt:

Mgr. Tomáš Kulhánek

CESNET z.s.p.o.

Zikova 4

16000 Praha 6

e-mail: tomaton@centrum.cz