

UNIVERZITA KARLOVA V PRAZE

FILOZOFICKÁ FAKULTA

Ústav informačních studií a knihovnictví

Bakalářská práce

Petr Laštovička

Využívání hlasových systémů a jejich uplatnění v informační společnosti

Voice systems usage and their application
in the information society

Praha 2016

Vedoucí bakalářské práce:
doc. PhDr. Richard Papík, PhD.

Motto: „*Opravdová moudrost je v poznání vlastní nevědomosti*“.

(Sokrates)

Na tomto místě bych rád poděkoval vedoucímu mé práce doc. PhDr. Richardu Papíkovi, PhD. za cenné rady ke koncepci práce a plánovanému dotazníkovému šetření.

Mé poděkování patří také prof. RNDr. Jiřímu Ivánkovi, za konzultaci k výslednému zhodnocení dat získaných v dotazníkovém průzkumu pro praktickou část této bakalářské práce.

Současně bych rád poděkoval PhDr. Evě Bratkové, PhD. za cenné konzultace k informacím z databázových informačních zdrojů, jejichž data jsem v této práci využil.

Prohlášení:

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

V Praze, dne 29. dubna 2016

.....

Petr Laštovička

Citace dle ISO 690:

LAŠTOVIČKA, Petr. Využívání hlasových systémů a jejich uplatnění v informační společnosti. 45s., 3s. příloh. Bakalářská práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví, 2016. Vedoucí bakalářské práce PhDr. Richard Papík, PhD.

Abstrakt:

Bakalářská práce se zabývá problematikou rozpoznání řeči a jejím praktickým využitím. Tato funkce je popsána teoreticky k více vědeckým oborům, stejně tak směrem k oboru Informační vědy s důrazem na tématickou oblast Human Computer Interaction.

Součástí práce je také dotazníkové šetření zaměřené na získání informací k chování uživatelů kontaktních center – jejich zvyklostí, frekvence využívání, preference typu obsluhy a požadavků na nové funkce. Obsahuje i specifické otázky k funkcím rozpoznání řeči a identifikaci hlasem. Jedná se o pilotní průzkum, který může být základem dalšího rozsáhlejšího výzkumu.

Cílem práce je zjistit možnosti využití funkcí rozpoznání řeči a reálné potřeby uživatelů.

Abstract:

This bachelor thesis deals with the problems of speech recognition and its practical use. This function is described theoretically for more scientific disciplines, as well as towards the Information Sciences with emphasis on the thematic area of Human Computer Interaction.

The work also includes a survey focused to obtain information on the behavior of users of contact centers - their habits, frequency of use, the type of service preference and request of new features. It also includes specific questions about the functions of speech recognition and voice identification;. This pilot survey can be the basis for further extensive research.

The aim is to determine the possibility of using speech recognition and the real needs of users.

Klíčová slova:

Rozpoznání řeči, Identifikace hlasem, Kontaktní centra, Call centra, HCI, Člověk, Počítač

Keywords:

Speech recognition, Voice identification, Contact centre, Call centre, HCI, Human, Computer

Obsah

Předmluva.....	7
1. Úvod	8
1.1. Technické obory a informační věda	9
1.2. HCI v kontextu s informační vědou	9
1.3. Četnost publikací / Informační zdroje	11
1.4. Rozpoznání hlasu ekonomicky	15
2. Teoretická část bakalářské práce.....	16
2.1. Hlasové technologie v praxi	16
2.2. Zpracování hlasu technicky.....	19
2.3. Kontaktní centra.....	22
2.4. Historie kontaktních center	23
2.5. Funkce kontaktních center	24
2.5.1. Příchozí hovory.....	24
2.5.2. Odchozí hovory.....	24
2.5.3. IVR - Interactive Voice Response	25
2.5.4. Nahrávání hovorů	25
2.5.5. Detekce emocí.....	25
2.6. Nové trendy	26
2.6.1. Identifikace hlasem	26
2.6.2. Ovládání hlasem.....	26
2.6.3. Inteligentní směrování příchozích hovorů.....	27
2.6.4. Automatizace odchozích hovorů	27
2.6.5. Převod textu na hlas a řeči na text.....	28
3. Praktická část bakalářské práce.....	29
3.1. Výběr metody	29
3.2. Metodika pilotního průzkumu	30
3.3. Volba průzkumného vzorku.....	30

3.4.	Definice výzkumné otázky a hypotézy	33
3.5.	Prezentace otázek a odpovědí.....	34
3.5.1.	Využívání služeb kontaktních center	36
3.5.2.	Způsob používání	37
3.5.3.	Volba funkčnosti hlasem	38
3.5.4.	Identifikace hlasem	39
3.6.	Vyhodnocení odpovědí	40
4.	Závěr	42
5.	Seznam použité literatury.....	43
6.	Seznam obrázků.....	45
Příloha I.	46

Předmluva

Jako téma své bakalářské práce jsem si vybral problematiku hlasových informačních systémů, ke které mám vzhledem k mému současnému zaměstnání blízký profesní vztah.

V rámci studia jsem získal řadu nových informací, které bych chtěl v rámci této práce využít a zúročit tak nabyté vědomosti. Kromě vlastního tématu hlasových systémů, které nás běžně provázejí v každodenním životě, bych rád prozkoumal jejich vazbu do moderní informační společnosti a způsob využívání nových nabízených možností.

Zajímavou kapitolou jsou také nové funkce a trendy v telekomunikačních systémech, kterými se v poslední době ubírají poskytovatelé těchto služeb v návaznosti na potřeby jejich uživatelů a zákazníků. Cílem je kromě zvýšení kvality poskytovaných služeb i zlepšení spokojenosti těchto uživatelů a to v rámci celého věkového spektra.

Není mou ambicí, aby tato bakalářská práce vyčerpávajícím způsobem popsala a vysvětlila kompletní problematiku hlasových systémů a jejich využití, tak jak je uvedeno v odborné literatuře. Naopak se hodlám zaměřit na zjištění detailní informace od stávajících uživatelů k provozovaným službám, kde se hlasové systémy již využívají.

V těchto systémech lze jistě nasadit nové funkce, například významněji spojené s rozpoznáním nebo identifikací hlasu, tak aby co nejvíce zjednodušily používání dané služby. Zbývá tedy jen zjistit, do jaké míry jsou tyto funkce ze strany uživatelů poptávány.

Bakalářská práce má celkem 78 266 znaků, 77 437 bez poznámkového aparátu, 45 stran.

K práci je připojena jedna příloha v rozsahu 3 stran, která se nachází za textem.

1. Úvod

Způsob komunikace mezi lidmi se v posledních dvaceti letech změnil v míře, kterou by nikdo pravděpodobně předtím neočekával. Moderní technologie využíváme každý den a mnohdy si již nedokážeme představit, že bychom je v běžném životě neměli dostupné.

Jednou z oblastí, kde lze tento posun v poslední době významně pozorovat, je oblast telekomunikací. Například využívání mobilních technologií umožňuje způsob práce a životního stylu, který by dříve vůbec nebyl možný. Moderní mobilní telefony tak již v současnosti disponují vestavěnou funkcí rozpoznání řeči a umožňují tak nový způsob zadávání vstupních informací a práce s tímto typem zařízení.

Rozvoj systémů pro rozpoznání hlasu, které zvláště v posledních deseti letech zažily nebývalý rozmach, se ukazuje jako aktuální trend, který může ovlivnit další oblasti života. Významně se to dotýká všech aktivit, kde se předpokládá rozhodnutí a ovládání hlasem je tím nejpřirozenějším způsobem. V současnosti dosáhla technologie takové úrovně, že kvalita rozpoznání hlasu se blíží úrovni odpovídající rozeznání informace druhou osobou. To umožňuje nasazení těchto systémů i v oblastech, kde bychom to v minulosti vůbec nepředpokládali.

Jednou z oblastí, kde se ve významné míře hlasové systémy uplatňují, je oblast telekomunikací. Zde například v telefonních kontaktních centrech, je novým trendem nasazení systémů pro rozpoznání hlasu volající osoby, které na jedné straně umožňuje jiný způsob zadání požadavku pro požadované funkce, na straně druhé lze hlasu volajícího využít pro jeho osobní identifikaci. Lze tak snížit množství nutných aktivit, které jsou v současnosti potřeba na aktuálních systémech, a navíc se zvýšenou bezpečností odpovídající identifikaci volajícího jeho vlastním hlasem.

Ve své práci tak chci popsat zkušenosti uživatelů s využíváním těchto telefonních systémů a odpovědět si na hypotetické otázky, které si kladu. Výzkumnou sondou tak uskutečňuji napříč věkovým spektrem se snahou porovnat odpovědi respondentů na různé otázky a to zvláště ve vztahu k současnému způsobu používání nebo i s možnostmi nově nabízených služeb.

Problematika hlasových systémů má celou řadu oblastí, kde lze nalézt spojitost s oborem Informační vědy. Zvláště v návaznosti na kontaktní centra se jedná o zpracování a poskytování informací v celém oboru lidské činnosti. V tomto případě se pak jedná o koncového uživatele, který s kontaktním centrem komunikuje různými kanály, jako je telefon, email, chat nebo web.

1.1. Technické obory a informační věda

Problematikou rozpoznání řeči se zabývají hlavně technické obory, v případě České republiky jsou to zvláště VUT v Brně, Technická univerzita v Liberci a ČVUT, katedra teorie obvodů. Poslední jmenovaná má s tímto typem výzkumu velké zkušenosti a již dlouhodobě zde probíhají semináře na multioborové bázi, kde se znalosti jednotlivých účastníků vzájemně doplňují. Příkladem přínosu našeho oboru do této problematiky jsou například uveřejněné případové studie „Bibliometrické a informetrické metody v databázích“, které detailně popisují jak efektivně pracovat v celosvětových databázích.

„Databázová centra vědeckých, technických a ekonomických informací se nemusí využívat jen k realizaci klasických rešeršních postupů, jejichž výsledkem jsou dokumenty sloužící pro vědeckou nebo odbornou práci, ale také ke speciálním analýzám, statistikám a výstupům, které vyplývají z bibliometrických a informetrických metod“ (Papík, 2004, s.169)

Obor Informačních studií je tak zde přínosem pro detailní znalost ověřených informačních zdrojů a celosvětových databází, z nichž lze čerpat podklady z již uskutečněných výzkumů.

1.2. HCI v kontextu s informační vědou

Human Computer Interaction (HCI) jako obor spojuje několik různých disciplín, kde se každá z nich zaměřuje na jinou část uživatelských rozhraní. Jednou z nich je ergonomie, která se od poloviny 20. století rozvinula do zkoumání vztahu člověka a strojů resp. možností jejich ovládní. Dalšími disciplínami jsou například lingvistika, design, informační věda, psychologie a sociologie, jejichž společným bodem je člověk ve vztahu k zařízení.

„Je zajímavé, že moderní informační a knihovní věda má velmi blízko k rozvoji ergonomických disciplín, které jsou základem právě pro obor HCI. Řada informačně-knihovnických škol ve světě má ve svých osnovách téma HCI zakomponováno a mnozí z učitelů jsou odborníky právě na tuto oblast - na člověka orientovaných disciplínách. Informační a knihovní věda se zabývá tzv. uživatelským a informačním chováním.“ (Papík, Souček, Drobíková, 2009, s.88)

Oblast Human Computer Interaction není zaměřena jen na komunikaci zrakem přijímaných informací, ale i na jiné smyslové úrovně. Příkladem může být oblast hlasové komunikace, která představuje způsob ovládní nejvíce přirozený koncovému uživateli. V současnosti se sice nejeví reálné, že by systémy či zdroje v digitálních knihovnách pracovaly na bázi hlasem ovládaných systémů, nicméně tato možnost je dnes již řešitelná, byť ekonomicky náročná.

Systémy rozpoznávání řeči, které jsou již dnes funkční, se dělí na systémy:

- rozpoznávání závislé na mluvčí osobě,
- adaptivní systémy přizpůsobivé mluvčí osobě,
- systémy rozpoznávání nezávislé na mluvčí osobě.

„Je pochopitelné, že trend se ubírá k systémům nezávislým na mluvčí osobě. Syntéza řeči a její rozpoznání má význam především v telefonním styku, kde nejsou k dispozici obrazovky, na nichž by se zobrazovaly informace dialogu. Hlasově ovládané uživatelské rozhraní může být kompatibilní s grafickými uživatelskými systémy. Je zřejmé, že při rozvíjení těchto rozhraní pro ovládání informačních systémů musí spolupracovat několik vědních oborů.“ (Papík, 2001)

Nicméně při nasazení systémů rozpoznání řeči lze shledat řadu nedostatků:

- hlas a dialekt každého člověka se liší,
- kvalita hlasu může být ovlivněna momentální situací (hluk, stres),
- kvalita mikrofону se významně podílí na dosaženém výsledku,
- hardware systému je náročnější na výpočetní výkon než u jiných systémů,
- systémy mohou mít omezenou slovní zásobu.

Nejsložitější etapou rozpoznávání řeči je tak porozumění jejímu obsahu. V této fázi musí dojít k přesné interpretaci toho, co chtěl mluvčí říct, tak aby výsledek jednoznačně odpovídal požadovanému sdělení. Je dobře známo, že porozumění řeči je založené na obrovském množství jazykových a kulturních znalostí.

Většina systémů pro rozpoznávání hlasu tak disponuje znalostmi o přirozeném jazyku.

Zde se pak významnou měrou uplatňují i další obory zaměřené na jazyk jako například:

Sémiotika - teorie znaků a znakových soustav, která zkoumá základní a složené znaky, v přirozených jazycích i formalizované znakové soustavy ze tří pohledů: syntax (struktura znaků), sémantika (vyjadřování smyslu znaků), pragmatika (vztah znaků a jejich uživatelů).

Syntaktika - sémiotická disciplína, která studuje jazyk z hlediska forem znaků, bez vztahu co jazykové výrazy znamenají. Zabývá se tedy pouze vztahy mezi znaky nebo řadami znaků, zkoumá vazby mezi slovy ve větě a správné tvoření větných konstrukcí a slovosledu.

Sémantika - součást sémiotiky zkoumající vztah mezi formou a významem znaku; v užším smyslu se odlišuje od sémantiky zkoumající vztah znaku ke skutečnosti.

Pragmatika - zabývá se řečovými akty a všímá si záměru i strategie mluvčího; člověk, který mluví, nejen něco říká, ale obvykle tím sleduje i nějaký záměr.

1.3. Četnost publikací / Informační zdroje

Téma rozpoznání řeči je v posledních letech velmi aktuální a ilustruje to množství uveřejněných prací, které se tímto tématem zabývají v rámci nějakého typu publikace. Při hledání lze zadáním klíčových slov zjistit rozsah takových publikovaných prací a to zvláště z informačních zdrojů, kde jsou taková klíčová slova k těmto dílům podle obsahu přidělena. Právě ověřená informace v daném elektronickém zdroji je důležitým faktorem, který potvrzuje relevantnost vyhledaného díla k danému tématu a současně poskytuje i detailnější informace o obsahu publikace.

Z hlediska mnou zkoumaného tématu jsou zajímavé zejména elektronické zdroje, které buď přímo cílí na technicky zaměřené databáze nebo na závěrečné práce, které se touto tematikou zabývají. V dostupných informačních kanálech, které jsou k dispozici v rámci univerzitní sítě, se jedná například o Portál elektronických zdrojů, zkráceně PEZ ¹. Ten obsahuje přímý přístup k velkému množství databází z mnoha oborů, které jsou předmětem studia na UK v Praze.

Ze seznamu takto dostupných elektronických zdrojů jsem vybral následující:

- Proquest

je multioborová databáze indexující tisíce odborných i populárně naučných časopisů, novin a dalších dokumentů. Je aktuálním nástupcem databází Dialog a Datastar, které nahradila od roku 2012. Poskytuje více než 22 tisíc odborných časopisů a plné texty více než 100 tisíc disertačních prací.

- Web Of Science

Společná platforma pro Web of Science Core Collection a Journal Citation Reports provozovaná společností Thomson Reuters. Obsah této databáze zahrnuje více než 10 000 nejvýznamnějších žurnálů z celého světa.

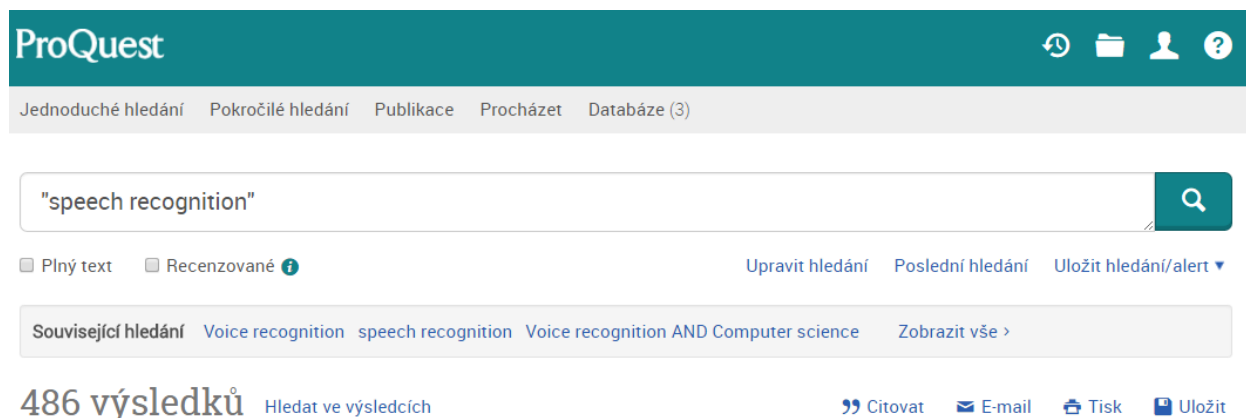
- IEEE Xplore, Digital Library

Jedná se o veřejnosti volně dostupnou digitální knihovnu profesní neziskové instituce Institute of Electrical and Electronics Engineers. Tento institut sdružuje nejvíce členů technické profese ve světě a to ve 150 zemích všech světadílů. IEEE vydává více než 100 titulů odborných periodik a pořádá konference.

¹ PEZ – <http://pez.cuni.cz/prehled/abecedne.php?lang=cs>

Jedna z nejznámějších multioborových databází **ProQuest** uvádí, že v časovém období posledních 25 let bylo k tomuto tématu uveřejněno více než 60 tisíc různých prací.

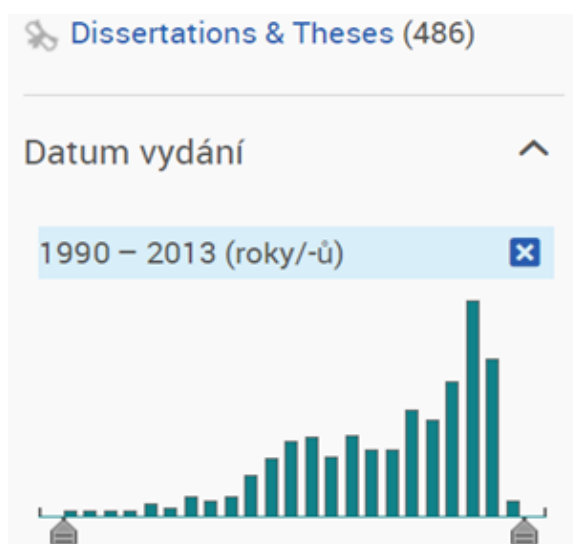
Pro vlastní vyhledání bylo použito klíčové slovo „speech recognition“, typ publikací byl omezen pouze na závěrečné práce (Dissertations & Theses) a rozsah času na roky 1990 – 2015. Při tomto zúženém hledání bylo nalezeno 486 prací na toto téma.



The screenshot shows the ProQuest search interface. At the top, the ProQuest logo is on the left, and navigation icons (refresh, folder, user, help) are on the right. Below the logo, there are links for 'Jednoduché hledání', 'Pokročilé hledání', 'Publikace', 'Procházet', and 'Databáze (3)'. A search bar contains the text '"speech recognition"' with a search icon on the right. Below the search bar, there are filters for 'Plný text' and 'Recenzované'. To the right of the filters are links for 'Upravit hledání', 'Poslední hledání', and 'Uložit hledání/alert'. Below the search bar, there is a section for 'Související hledání' with links for 'Voice recognition', 'speech recognition', and 'Voice recognition AND Computer science', and a 'Zobrazit vše >' link. At the bottom of the search results, it says '486 výsledků' and 'Hledat ve výsledcích'. To the right of this are icons for 'Citovat', 'E-mail', 'Tisk', and 'Uložit'.

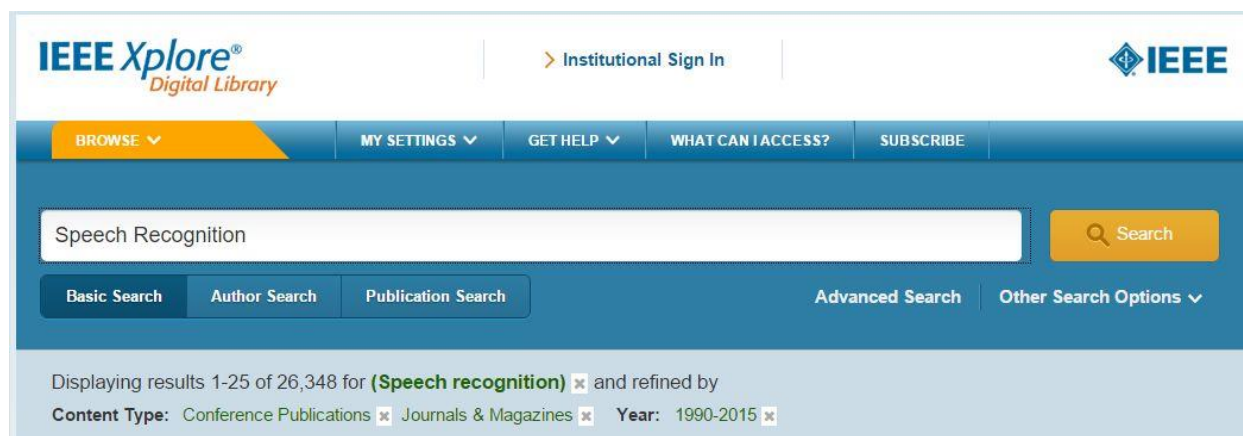
Obrázek 1: Uživatelské rozhraní Proquest

Četnost závěrečných prací s tématem rozpoznání řeči je graficky vyjádřena níže a názorně zobrazuje růst množství prací v čase, zřejmý obzvláště v posledních letech. Tato informace tak potvrzuje rostoucí zájem o problematiku rozpoznání řeči jako tématu studia.



Obrázek 2: Zobrazení vyhledaných prací v čase
(Proquest, 2016)

Dalším z velmi známých informačních zdrojů a to digitální archiv neziskové organizace **IEEE**, uvádí ke klíčovému slovu „Speech Recognition“ více než 26.000 tisíc odkazů:

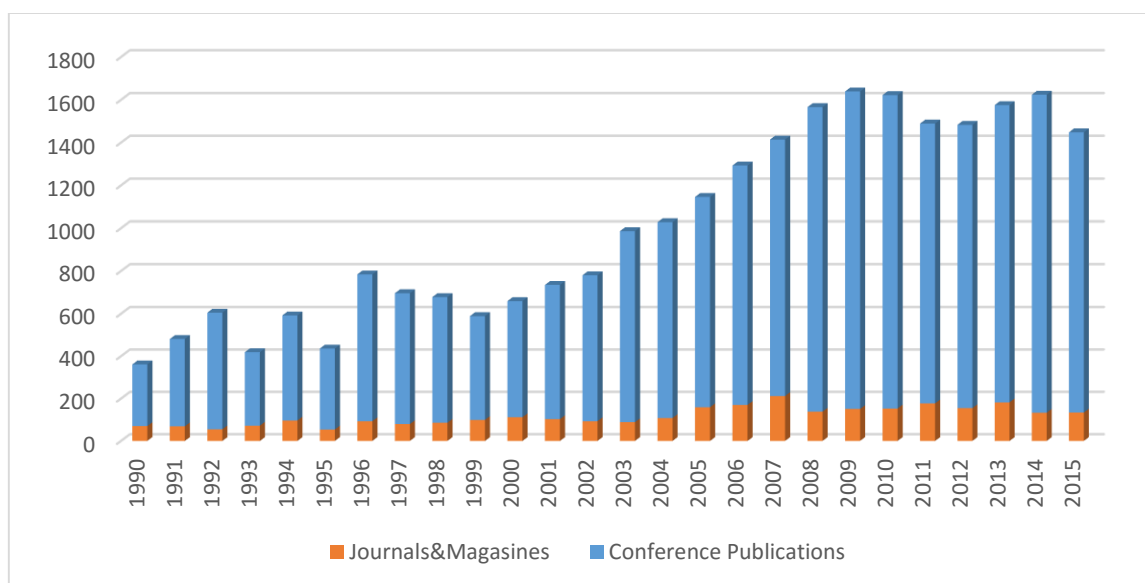


Obrázek 3: Uživatelské rozhraní IEEE Xplore

Ve vybraném časovém rozsahu let 1990 až 2015 je v jednotlivých rocích znatelný nárůst a to jak v počtu publikací z konferencí, tak i nalezeném počtu pro časopisy:

IEEE	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Součet
Conference Publications	291	412	550	447	496	382	691	616	592	490	547	632	787	897	920	987	1123	1202	1428	1488	1469	1312	1328	1393	1491	1314	23285
Journals&Magasines	70	69	55	72	96	54	93	80	86	99	112	103	113	89	108	159	170	212	138	151	153	177	155	182	133	134	3063

V grafickém vyjádření je rostoucí trend počtu hodnot ještě zřetelnější:



Obrázek 4: Zobrazení vyhledaných prací v čase

(IEEE, 2016)

Dalším z významných informačních zdrojů v oblasti technických oborů je soubor 7 databází provozovaných společností Thomson Reuters a známý pod jménem **Web of Science**.

Ke klíčovému slovu „Speech Recognition“ zobrazil v posledním čtvrtstoletí 19.006 odkazů:

The screenshot shows the Web of Science search results for the query "speech recognition". The results are sorted by "Publication Date -- newest to oldest". The page shows 19,006 results from the Web of Science Core Collection. The search was conducted between 1990 and 2015. The interface includes a search bar, navigation buttons like "My Tools", "Search History", and "Marked List", and options to "Save to EndNote online" and "Add to Marked List".

Obrázek 5: Uživatelské rozhraní Web Of Science

Pokud ještě více zúžíme výběr v čase, například na posledních 8 let, počet vyhledaných publikací klesne na číslo 9.677. Rozhraní umožňuje i zajímavou funkci a to stiskem odkazu "Citation Report" zobrazit počet citací v jednotlivých letech a u jednotlivých publikací v roce.

The screenshot displays the "Citation Report: 9677" for the search term "speech recognition". It features two bar charts: "Published Items in Each Year" and "Citations in Each Year". The "Published Items in Each Year" chart shows a steady decline from approximately 1300 items in 2007 to 1200 in 2014. The "Citations in Each Year" chart shows a significant increase from about 1,000 citations in 2007 to nearly 8,000 in 2015. To the right of the charts, a summary table provides key statistics:

Results found:	9677
Sum of the Times Cited [?]:	38879
Sum of Times Cited without self-citations [?]:	27977
Citing Articles [?]:	21312
Citing Articles without self-citations [?]:	17235
Average Citations per Item [?]:	4.02
h-index [?]:	61

Below the charts, there is a table showing the citation data for individual items, sorted by "Times Cited -- highest to lowest". The table includes columns for the years 2012 through 2016, a "Total" column, and an "Average Citations per Year" column. Two items are listed:

	2012	2013	2014	2015	2016	Total	Average Citations per Year
1. Deep Neural Networks for Acoustic Modeling in Speech Recognition By: Hinton, Geoffrey; Deng, Li; Yu, Dong; et al. IEEE SIGNAL PROCESSING MAGAZINE Volume: 29 Issue: 6 Pages: 82-97 Published: NOV 2012	1	59	99	109	30	298	59.60
2. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition By: Dahl, George E.; Yu, Dong; Deng, Li; et al. IEEE TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING Volume: 20 Issue: 1 Pages: 30-42	14	58	90	80	21	263	52.60

Obrázek 6 : Zobrazení citací vyhledaných prací

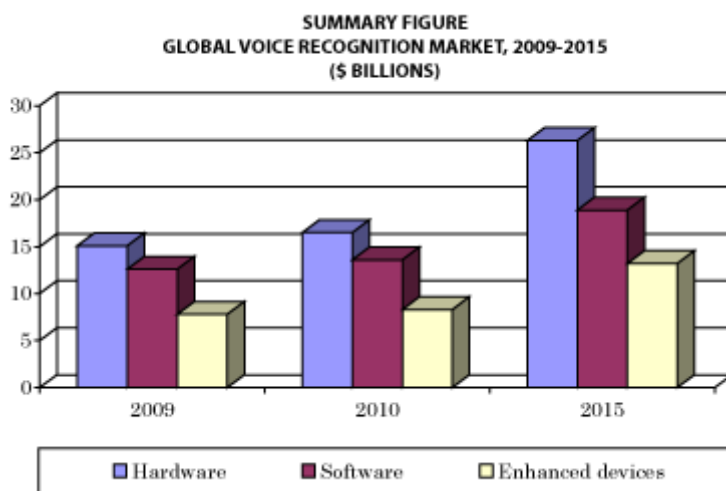
(Web Of Science, 2016)

1.4. Rozpoznání hlasu ekonomicky

Firmy napříč všemi sektory hledají svoji konkurenční výhodu, která by je vzájemně odlišila ve stále přeplněnějším podnikatelském prostředí. Chtějí tak produkty, které jim mohou nejen udržet jejich zákazníky, ale současně pomohou k růstu jejich podnikání. To se například týká makléřských firem, leteckých společností, bank nebo pojišťoven, které v budoucnu spoléhají na rozvoj funkcí k rozpoznávání hlasu, aby tak zlepšily své kontakty se zákazníky.

To, že tento sektor je i z ekonomického pohledu zajímavý, vyplývá z výsledků výzkumu společnosti BCC Search, která zkoumala tuto část trhu v období let 2009 až 2011 s výhledem do budoucna na dalších 5 let. Výsledky výzkumu jsou shrnuty v následujících bodech:

- *Trh s technologiemi pro rozpoznání hlasu má aktuální růst v úrovni 8.8 % a to i do období let 2010 a 2015. Celkový trh je hodnocen přibližnou částkou \$38.4 miliard dolarů v roce 2010 s očekávaným nárůstem na \$58.4 miliard dolarů v roce 2015.*
- *Software pro rozpoznání hlasu vyžaduje speciální hardware a to nejen k přenosu signálů, ale i k eliminaci okolního šumu a vyhodnocení komplexních podmínek zvuku. Tento sektor trhu se v roce 2010 pohyboval na úrovni \$16.5 miliard dolarů s růstem 9.8 % a očekávaným objemem trhu v roce 2015 na úrovni \$26.3 miliard dolarů.*
- *Pokud software pro rozpoznání mluvené řeči (Automatic speech recognition) a syntézu řeči (Text-to-Speech) pracují společně, pak umožňují nasazení aplikací ovládaných hlasem. Prodej takových typů software roste o přibližně 6.8 % ročně a to z hodnoty \$13.6 v roce 2010 na očekávanou úroveň \$18.9 miliard dolarů v roce 2015.*



Obrázek 7: Trh s aplikacemi pro rozpoznání řeči

(BCC Research, 2010)

2. Teoretická část bakalářské práce

V této části práce bych rád uvedl základní informace, které se pojí s problematikou rozpoznání řeči. Kromě technického zařízení, které se na převodu hlasu významně podílí, je ale nedílnou součástí celého systému i znalost řeči tvořená jednotlivými slovy a jejich vzájemnou vazbou.

Důležitým předpokladem je porozumění skladbě věty, kde existují přísná sémantická pravidla, podle kterých je nutné se řídit, aby byl systém schopen kombinaci slov ve větě pochopit. Stále je třeba pokračovat v důkladném výzkumu a dále pracovat s takovými vlastnostmi řeči, jako je morfologie, důraz, výška zvuku, rychlost, hlasitost, slučování slov, kontext, artikulace, synonyma nebo lingvistická informace. Předpokládá se, že hlavním směrem vývoje bude modelování jazyků vhodně použitelné v systémech rozpoznávání řeči. Důraz je kladen na integrovaný soubor gramatických pravidel, kde čím přesnější pravidla jsou, tím snazší je systém realizovat.

Této problematice se například aktuálně věnuje práce kolektivu autorů Skarnitzl, Šturm a Volín, kteří publikovali dílo s názvem *Zvuková báze řečové komunikace: fonetický a fonologický popis řeči*, jež cílí na specifika analýzy řeči v českém jazyce.

2.1. Hlasové technologie v praxi

Snadnost s jakou člověk zvládá mluvení a naslouchání vedla k myšlence vybavit počítače (a počítačem řízené stroje) novým rozhraním, které by umožňovalo komunikovat s nimi prostřednictvím hlasu. Hlavním úkolem řečového rozhraní je porozumět lidské řeči a adekvátně reagovat na hlasové příkazy. Úkolem je také naučit jednotky rozumět řeči bez „prostředníka jazyka“, ve kterém lidé komunikují mezi sebou, což znamená, že je potřeba vymyslet algoritmus pro rozpoznání zvukového signálu řeči. Přesně toto je účelem technologie rozpoznávání řeči.

„Hlasový signál, jehož původcem je člověk, je v porovnání se syntetickým signálem značně nedeterministický, vysoce variabilní v závislosti na konkrétní osobě, na jejím momentálním fyzickém i psychickém stavu a na jejích vyjadřovacích schopnostech, navíc je ještě ovlivněný okolním prostředím. Takto složitý signál, v němž informace o vlastním obsahu tvoří jen malou složku, má být analyzován počítačem, jehož možnosti jsou v porovnání s mozkem člověka stále ještě velmi omezené. Místo porozumění obsahu řeči tak počítač pouze více či méně úspěšně dekóduje některé úseky, převádí je na předem stanovenou množinu textových symbolů, případně na jejich základě vykonává určité akce.“ (Nouza, 2011, s.152)

Čím více informací víme o vstupním signálu, tím snadněji dokážeme zpracovat a rozpoznat tuto informaci. Pokud je slovník příliš malý, každé slovo je vysloveno zvlášť a jedná se o tutéž osobu, lze použít algoritmy, které sekvence příznakových vektorů porovnávají s referenčními sekvencemi odpovídajícím slovům, které namluvila tatáž osoba. Složitější je pak systém, který má sloužit neomezenému okruhu uživatelů, a to okamžitě bez předchozí adaptace. Používá-li se nepřilíš omezený slovník v rozsahu několika desítek slov, pak tyto systémy pracují tím spolehlivěji, čím více osob poskytlo nahrávky svých promluv.

Systémy založené na těchto pravděpodobnostních modelech – nejčastěji se používají „skryté markovské modely“² – využívají tzv. trénování systému, které je časově i výpočetně mnohem složitější než vlastní rozpoznávání. Provádí se však pouze jednou v době vývoje dané aplikace a vytvořené modely v praxi výrazně urychlují vlastní rozpoznávání probíhající v reálném čase.

Markovský skrytý model umožňuje určit pravděpodobnost výstupních hodnot nezávisle na posloupnosti vnitřních stavů systému, kterými model prošel. Vnitřní stavy systému, na kterých pak závisí stav výstupu, nejsou pozorovateli viditelné, a proto se používá přívlastku „skryté“.

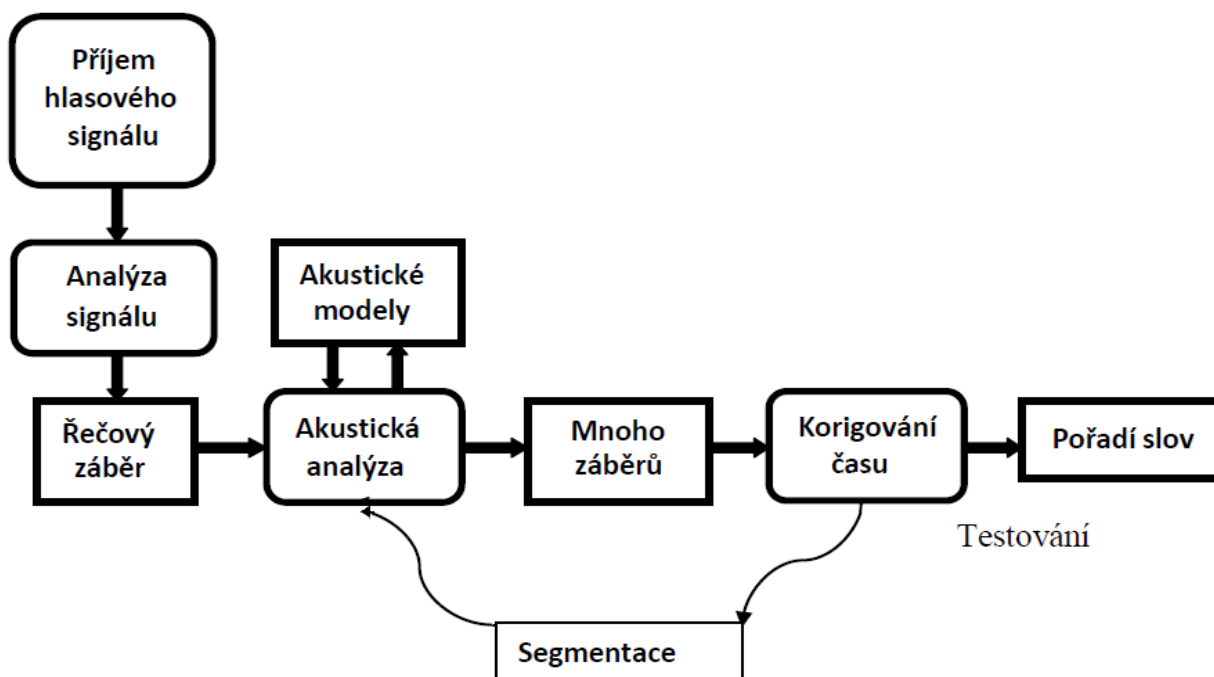
„Z několika praktických důvodů již není výhodné pracovat s celými slovy, ale jít směrem k nižším řečovým jednotkám. Těmi nevhodnějšími pro rozpoznání řeči jsou pak hlásky neboli „fonémy“, jejichž počet je totiž poměrně malý. Ve většině evropských jazyků se pohybuje v rozsahu 30 až 60 a lze z nich sestavit libovolné slovo v daném jazyce. Aby tyto modely dobře reprezentovaly jednotlivé hlásky, musí být k dispozici mnoho variant záznamů promluv s jejich fonetickou anotací a přesným určením, kde se v záznamech nacházejí.“ (Nouza, 2011, s.154)

Získávání hlasového signálu neboli diskretizace hlasu je definovaná jako proces zpracování zvukového signálu, kde existují dva typy zvuků - znělé a neznělé. Znělé zvuky tvoří vibrace hlasivek při průchodu vzduchu a jejich rezonanční signál se skládá ze čtyř frekvenčních složek, tzv. formantů, které jsou hlasovým razítkem pro zvuky produkované lidským řečovým aparátem. Neznělé zvuky pak produkuje vzduch procházející hlasovým kanálem bez vibrací hlasivek.

Oba typy zvuků jsou považované za časové řady dat, převzaté z pravidelných časových intervalů, kde se pro izolaci požadovaného intervalu používají prostorová okna. Některé funkce okna vypočítávají průměrnou amplitudu, jiné pak počet nulových přechodů a transformace „Fourierova signálu“³ během intervalu. Pro odstranění hluku se používají různé filtrační techniky.

² Markovovy řetězce - statistická metoda objevená ruským matematikem A. A. Markovem (1856-1922)

³ Fourierova transformace - slouží pro převod signálů z časové oblasti do oblasti frekvenční



Obrázek 8: Etapy rozpoznávání řeči (Yemchenko, 2011)

Rozpoznávání řeči tak představuje víceúrovňový úkol rozpoznávání obrazů, ve kterém jsou akustické signály analyzovány a strukturovány do hierarchie konstrukčních prvků (např. fonémy, slova, fráze a věty). Vlastní proces rozpoznávání hlasu probíhá v několika fázích, které lze rozdělit do několika hlavních etap názorně zobrazených ve schématu - Etapy rozpoznávání řeči.

První etapou je příjem hlasového signálu a předběžné zpracování řeči. V druhé etapě jde o rozpoznávání fonémů a slov. Třetí závěrečnou etapou je vlastní porozumění řeči, které je nejsložitější částí celého procesu rozpoznávání řeči. V této fázi musí být pořadí slov ve větě proměněné na přesnou reprezentaci toho, co chtěl mluvčí ve skutečnosti sdělit. „Rozumět“ řeči je totiž to nejtěžší. (Yemchenko, 2011)

Na cestě rozvoje rozpoznávání řeči stojí několik zásadních překážek, například nedostatečný objem slovníků, vytvoření šablon nepřetržité řeči, různé typy přízvuků a výslovnosti. Objemy slovníků definují stupeň složitosti, požadavky na výpočetní výkon a spolehlivost systémů. Je zřejmé, že čím větší je obsah slovníku zapsaného do systému rozpoznávání, tím větší je četnost chyb při rozpoznávání slov systémem. Například slovník, který se skládá z 10 číslic, může být rozpoznáván prakticky bezchybně, zatímco četnost chyb při rozpoznávání slovníku velikosti 100 000 slov může dosáhnout vyšší chybovosti. Na druhou stranu může rozpoznávání malého slovníku produkovat velké množství chyb, pokud jsou si slova navzájem velmi podobná. Důležitým bodem je také účel systému rozpoznávání řeči, který určuje požadovanou úroveň abstrakce, na kterém se bude provádět rozpoznávání řeči.

2.2. Zpracování hlasu technicky

Pro zpracování hlasu v počítačových aplikacích je nejprve nutné vlastní hlas digitalizovat, tak aby s ním bylo možné nadále pracovat. Jakákoliv informace, která má být zaznamenána nebo přenesena komunikačními systémy, musí být nejdříve převedena na elektrický signál.

K tomuto účelu slouží převodníky zvukových veličin na veličiny elektrické, nejčastěji napětí. Vytvořený elektrický signál, obvykle v analogové podobě, se následně převádí na signál digitální a to v analogově/digitálním převodníku.

Řečový signál má sice svá specifika, ale lze na něj aplikovat i obecné metody zpracování signálů. Stále se hledají a vyvíjí specializované metody pro zpracování řečového signálu, které umožňují vyšší kompresi a tím snižují požadavky na kapacitu přenosového kanálu. Klasickými metodami zpracování signálu aplikovanými na řečový signál však není možné dosáhnout tak výrazné úspory jako metodami specializovanými. Úroveň komprese se vyjadřuje veličinou zvanou kompresní poměr, což je vztah mezi vstupní a výstupní velikostí souboru.

Pro zdrojové kódování řečových signálů (do maximální frekvence 4 kHz) se používají kodéry, které je možné dělit do tří základních skupin:

- Kodéry tvarového průběhu (Waveform Coders)
Realizují tzv. zdrojové kódování tvaru vlny a jsou konstruovány tak, aby se časový průběh analogového signálu na výstupu dekodéru co nejvíce shodoval s časovým průběhem analogového signálu na vstupu kodéru.
- Vokodéry (Vocoders, Voice Coders)
Realizují parametrické zdrojové kódování, kde není přenášen původní signál, ale pouze jeho charakteristické parametry, vytvořené na základě analýzy původního signálu. Na přijímací straně je pomocí těchto parametrů řízen syntezátor hovorových signálů. Reprodukovaný signál má výrazný syntetický charakter.
- Kodéry hybridní
Vznikají kombinací kodérů obou předchozích skupin a vhodným způsobem spojují jejich přednosti. Realizují hybridní zdrojové kódování.

Nejjednodušší metoda převodu analogového signálu na signál digitální se nazývá PCM - Pulse Code Modulation a byla vytvořena již v roce 1938. Převod vstupního zvuku na digitální signál probíhá při této metodě ve třech krocích: vzorkování, kvantování a kódování.

Vzorkování odebrává v reálném čase vzorek ze spojitého signálu a to s maximální možnou frekvencí, kterou je možno navzorkovat. Proto je ještě před vzorkováním signál frekvenčně omezen dolní propustí. Při kvantování může každý odebraný vzorek nabývat hodnoty z oboru reálných čísel, proto je nutné určit diskrétní úrovně, ke kterým bude hodnota vzorku zaokrouhlena. Kódování přiřadí ke každému kvantovanému vzorku binární kód, kódové slovo.

„Pro analogový řečový signál v obvyklém rozsahu 300-3400 Hz, je podle vzorkovacího teorému při této metodě zvolena vzorkovací frekvence 8 kHz. Počet kvantovacích hladin je pak určen dynamikou signálu a hodnota každého vzorku je zaokrouhlena na jednu z 256 kvantovacích hladin. Tento postup popisuje lineární kvantování. Aby byla zajištěna rovnoměrná hodnota kvantizačního zkreslení v celém pracovním rozsahu kodéru, je vhodné použít nelineárního rozložení kvantizačních stupňů, která spočívá ve změně dynamiky signálu. Na vysílací straně je použit kompresor, tzn. že malé hodnoty vzorků jsou zesíleny a velké hodnoty vzorků zeslabeny. V přijímači je pak použit expandor, ve kterém probíhá inverzní proces.“ (Leitner, 2009)

V lokálním zařízení, kterým může být počítač, tablet nebo mobilní telefon, je vstupní signál k další analýze zprostředkován přímo mikrofonem napojeným na vstupní obvody dekodéru a je k dispozici v celém zvukovém spektru omezeném pouze kvalitou použitého mikrofonu. Analýza řeči tak probíhá nad zvukem v celém slyšitelném pásmu, který může být ovlivněn přidanými veličinami jako například šum, hluk na pozadí a podobně.

Aby hlas jednoho účastníka telefonního hovoru doputoval přes komunikační síť k druhému, musí být jistým způsobem zpracován a přenesen. V současné době již tedy nejde o klasický telefonní přenos, ale přenášená informace prochází několika různými druhy sítí a přenášená data mohou být tímto přenosem výrazně ovlivněna. V klasických telefonních sítích má na výsledek zpracování hlasu vliv hlavně kvalita telefonního přístroje, akustické echo, interference mezi linkami, šum a přeslechy. V mobilních sítích hraje hlavní roli úroveň signálu v radiovém prostředí, který má přímý vliv na data přenášená mezi mobilní stanicí a základnovou stanicí. Chyby v komunikačním kanále, které mohou být způsobeny frekvenčními interferencemi nebo impulzním ručením, je třeba řešit použitím opravných kódů. Dalším typem přenosové cesty jsou paketové sítě, dnes významně používané v souvislosti s IP telefoníí. V těchto typech sítí se objevují problémy vyplývající z principu funkce toho přenosu a to především ztrátovosti paketů, zpoždění doručení paketů nebo správné pořadí přenosu paketů.

Na přenášený hlas má významný dopad i typ použitého kodeku, kde typ použitého algoritmu více či méně ovlivňuje přenášený signál. Důležitá je také šířka hovorového pásma, kterou je schopen daný kodér zpracovat, dle kterého se kodeky dělí na úzkopásmové (signál do 3,5 kHz) a širokopásmové (do 7 kHz). Dalším sledovaným parametrem je zpoždění signálu způsobené převodem původně analogového signálu na digitální a zpět. Většinou se udává v hodnotě počtu milisekund, s tím že čím nižší je tím lépe, ale řádově by nemělo přesáhnout 100 ms.

Jedním z nejčastěji využívaných typů kodeku je standard ITU-G.711⁴, který je základním kamenem kódování hlasového signálu v klasických digitálních telefonních sítích. Tato norma definuje dva typy kompresní charakteristiky a to A-law, používanou v Evropě a Austrálii a μ -law, používanou v Americe a Japonsku. Rozdíl mezi oběma charakteristikami je v kompresním poměru, kdy A-law převádí vstupní signál na 8 bitové vzorky, zatímco μ -law na vzorky 7 bitové. Tento rozdíl je dán odlišnou strukturou telefonních sítí v Evropě a Americe.

„Princip komprese spočívá v nerovnoměrném rozložení kvantovacích hladin, jejich četnost klesá s rostoucími hodnotami úrovní signálu. Vychází se totiž z charakteru řečového signálu, kdy se předpokládá menší pravděpodobnost výskytu vysokých hodnot úrovně řečového signálu.

Původní 16 bitové vzorky jsou oříznuty na 13 bitové, které jsou pak převedeny danou kompresní charakteristikou na 8 bitové. V Evropě je pro přenos hlasu určených právě 8 bitů a signalizace se přenáší ve zvláštním kanále. Základní linka obsahuje 32 kanálů.

V Americe je pro přenos hlasu určeno jen 7 bitů, osmý je používán jako signalizační a základní linka obsahuje jen 24 kanálů. Pokud dojde k použití jakékoliv komprese, dá se předpokládat, že na druhé straně přenosového kanálu je nutné použít dekompresi. Na straně příjemce je použita inverzní charakteristika a dochází k expanzi signálu do původního průběhu.“ (Leitner, 2009)

Rozpoznání řeči při telefonních hovorech je v mnoha ohledech ovlivněno dalšími vlivy, které se vyskytují jak při přenosu signálu od zdroje k příjemci, tak zkreslením nebo výpadky informací podle typu přenosové sítě. Navíc se může významněji projevit užší hovorové pásmo dané šířkou frekvenčního spektra, ve které přenos hlasu probíhá. Jedním z dalších vlivů je možné zkreslení výstupního signálu oproti původnímu vstupnímu signálu právě typem použitého kodeku. Čím nižší bitová rychlost, tím lépe pro přenosový kanál, ovšem ne vždy pro kvalitu rekonstruovaného řečového signálu. To může mít vliv na kvalitu rozpoznání hlasu, přenášeného po komunikačních sítích, a to zda byl digitalizovaný hlas přenesen klasickou telefonní sítí, mobilní nebo paketovou sítí s různou mírou vlivu této sítě na přenášený signál.

⁴ ITU - *International Telecommunication Union* - Specializovaná agentura OSN se zabývá standardizací a tvorbou doporučení pro informační a komunikační technologie.

2.3. Kontaktní centra

Dnešní technologie umožňují, například v telefonních kontaktních centrech, rozšířit původní, čistě telefonní část, o moderní nadstavby ve formě automatizovaných systémů, nově také doplněné o moduly rozpoznání řeči. Mnohdy tak plně nahradí obsluhu a dodání požadované informace je rychlejší a na lepší úrovni než dříve. U části uživatelů však používání těchto služeb může působit specifické problémy, protože na jednu stranu nabízejí větší interaktivitu, ale na druhou stranu mohou být špatně použitelné, právě pro svou přílišnou modernost.

Otázkou je, co je vlastně přesně myšleno názvem „kontaktní centrum“⁵ respektive „call centrum“? Tento pojem nejlépe vystihuje práce k tématu kontaktních center (Smolová, 2011), která definuje prostředí kontaktních center a popisuje ho jako:

„Pracovní prostředí, kde je hlavní činnost prováděna prostřednictvím telefonu v součinnosti se zobrazovacím zařízením“. Call operátoři jsou definováni jako „osoby, jejichž práce vyžaduje, aby strávily převážnou část své pracovní doby reagováním na telefonáty, zatímco současně používají zobrazovací zařízení“ (Sprigg, Smith, Jackson, 2003).⁶

Zatímco dříve byla aktivita obchodních společností v oblasti kontaktních center považována za okrajovou oblast zájmu, postupem času se identifikoval jejich přínos nejen pro firmy, které tak mohly ovlivňovat své klienty, ale i pro zákazníky samotné.

V současnosti je oblast kontaktních center definována jako velmi klíčová a na kvalitu všech poskytovaných služeb je kladen velký důraz (Kociánová, 2007).

„Z původních call center se vyvíjejí kontaktní centra, která kromě telefonní komunikace využívají ke kontaktu se svými popř. potenciálními klienty i dalších prostředků jako je email, fax, SMS zprávy, webové formuláře, apod. Zatímco před lety byla hlavním impulsem pro klienta cena výrobku či služby, v dnešní době se stále častěji projevuje preference vstřícnější komunikace a lepší interakce mezi zákazníkem a podnikem před výší ceny. Kontaktní centra dnes již hrají klíčovou roli ve strategii podniku, ve velké míře totiž utvářejí vztah klienta k dané společnosti.“

U některých typů společností dokonce tvoří činnost kontaktního centra jeho hlavní obchodní model, protože nabízejí jednotlivé funkce svého kontaktního centra jiným institucím nebo obchodním společnostem formou komerční služby.

⁵ Řeší všechny typy kanálů včetně SMS, faxů, emailu a chatu.

⁶ <http://www.hse.gov.uk/research/rrpdf/rr169.pdf>

2.4. Historie kontaktních center

Vznik telefonních call center, tak jak je známe dnes, má svůj původ téměř před čtyřiceti lety a to jako služba pro rezervaci letenek společnosti Continental Airlines. Tento systém distribuce telefonních hovorů byl vyvinut americkou firmou US Rockwell, které se tak stala prvním úspěšným dodavatelem těchto systémů pro oblast obchodu a služeb. S rozvojem privátních telefonních systémů instalovaných přímo ve firmách se následně stala funkce telefonního call centra integrální součástí těchto systémů, aby časem nabyla na významu jako jeden ze základních marketingových znaků vstřícného vztahu k zákazníkovi.

V osmdesátých letech začala vznikat komerční telefonní call centra, jejichž obchodní model byl založen na prodeji po telefonu, zpočátku zvláště produktů typu pojištění a bankovních služeb. Ke stejné době se datuje vznik tzv. barevných linek, které zákazníkům umožnily volat zdarma nebo se sníženým tarifem za telefonní hovor. Kontaktní centra se tak stala jedním ze základních prvků komunikace se zákazníkem a to pro všechny typy nabízených služeb.

S rostoucím provozem a množstvím odbavených hovorů se v průběhu využívání začala řešit ekonomicky výhodnější řešení a to jak z hlediska počtu pracovníků poskytujících služby, tak současně z hlediska spokojenosti zákazníků, kteří do kontaktních center volali.

Díky významnému rozvoji komunikační techniky v devadesátých letech bylo možné připojit k telefonnímu systému další přídavná zařízení, která umožnila rozšíření používaných funkcí. Zvláště systémy automatické obsluhy, které bylo možné ovládat tónovou volbou pomocí stisku čísel na telefonu, byly významným krokem vpřed v zapojení volajícího do procesu hovoru.

Tyto systémy, pro které se i v Česku vžilo pojmenování IVR (Interactive Voice Response), tak umožnily volajícímu aktivně ovlivnit průběh hovoru nebo dokonce přímo získat požadovanou informaci bez kontaktu s živým pracovníkem kontaktního centra. Další podstatnou výhodou, které nasazení těchto systémů přineslo, byla jejich plná funkčnost 24 hodin denně.

Dalším faktorem, kterým se zvýšila efektivita, a dále rozšířila možnosti kontaktních center, bylo propojení telefonních systémů s externími počítačovými aplikacemi pomocí protokolu CSTA⁷. Ten je standardizovaný v systémech různých výrobců a umožňuje tak podporu integrací stejné aplikace k různým typům zařízení. Tento protokol je průběžně aktualizován o nové normy a reaguje tak na rozvoj komunikačních platforem, jejich funkcí a možných koncových přístrojů.

Průběžně byly také do kontaktních center integrovány i další komunikační kanály jako je fax, email, chat nebo propojení přes webové rozhraní například do sociálních sítí.

⁷ CSTA – Computer supported telecommunications application

2.5. Funkce kontaktních center

Hlavní funkcí kontaktního centra a zvláště jeho telefonní části je co nejlépe poskytnout službu zákazníkovi, který buď sám volá na informační linku nebo je mu aktivně z call centra voláno.

V případě zákazníka volajícího na informační linku se tak nejčastěji jedná o určitou formu získání informace, například stavu účtu, objednávky nebo průběh platby případně o vyřízení nějakého konkrétního požadavku - aktivace služby, provedení objednávky, reklamace, apod. Pokud je naopak zákazníkovi voláno, jedná se zpravidla o některou z variant tele-marketingu, jako například telefonický průzkum, obchodní nabídka nebo informační sdělení k již existující službě nebo produktu.

2.5.1. Příchozí hovory

Běžný příchozí hovor je směřován na vybraného pracovníka kontaktního centra podle předem stanovených kritérií, tak aby volajícímu co nejlépe sdělil informaci, kterou požaduje. Směrování hovoru může proběhnout na základě mnoha kritérií, kterými mohou být například:

- číslo volané služby, které určuje chování informační linky a její nabídku,
- číslo volajícího, jež lze využít k identifikaci podle údajů v databázi klientů,
- den a čas volání, kdy se linka může chovat rozdílně podle denní doby a dnu v týdnu.

Dalším důležitým faktorem je také zvolený jazyk komunikace, který kromě správných hlásek call centra určuje i jazyk, kterým musí hovořit pracovník, na kterého bude takový hovor přepojen.

2.5.2. Odchozí hovory

Nejčastější aktivitou, kdy pracovníci kontaktních center volají aktivně na známá telefonní čísla, jsou zpravidla činnosti obchodního charakteru, jako například telemarketing nebo kampaně, kdy probíhá volání klientům s aktuálními obchodními nabídkami nebo informacemi.

Další častou funkcí kontaktních center jsou také odchozí telefonní volání, kdy volající sám aktivně požádal o volání zpět na své číslo. Jedná se zejména o:

- volání klientům, kteří požádali o zavolání jiným kanálem (email, web nebo chat),
- volání na číslo, které si vyžádalo zpětné zavolání, např. při čekání ve frontě hovorů,
- volání na číslo, které volalo mimo pracovní dobu kontaktního centra.

Výhodou tohoto typu volání je možnost tyto odchozí hovory plánovat, což na jedné straně vytváří specializovanou činnost části kontaktního centra nebo na straně druhé umožňuje využít pracovníky kontaktního centra, kteří právě neodbavují žádný příchozí hovor a jsou neobsazení.

2.5.3. IVR - Interactive Voice Response

Automatický hlasový systém, pro který se i v Česku vžil původní anglický název IVR, umožňuje zákazníkům získat vybrané informace zpravidla pomocí stisku tlačítek na klávesnici telefonu s tónovou volbou. Hlavní výhodou IVR systému je, že je plně dostupný v kteroukoli denní dobu, 365 dní v roce / 7 dní v týdnu / 24 hodin denně bez přítomnosti obsluhy. Ve většině takových hlasových systému se setkáváme s těmito základními funkcemi:

- výběr služby z nabídky pomocí volby čísla,
- identifikace klientským číslem a heslem,
- hlasové sdělení informace, například stavu účtu nebo jiných číselných údajů,
- možnost zadat číslo mobilu.

Kromě vlastních funkcí se přímo nabízí možnost integrovat IVR systém s počítačovými systémy a databázemi zákazníka nebo nadstavbovými aplikacemi pro zpracování hlasu.

2.5.4. Nahrávání hovorů

Další funkcí, v některých případech dokonce ze zákona povinnou, je nahrávání proběhlých hovorů, které jsou následně po stanovanou dobu uloženy. Ke každé nahrávce je kromě datumu a času uskutečnění hovoru, možné připojit i další informace, které se ke konkrétnímu hovoru přidaly v rámci jeho směrování, zpracování a výsledném hovoru v daném kontaktním centru. Kromě vlastní archivační funkce, pomocí které je možné přehrání hovoru v budoucnu, je také možnost tyto nahrávky analyzovat. Informace z nahrávky lze využít, jak ke zkvalitnění služeb daného centra hovorů, tak k analýze obsahu hovoru. Tento obsah může být následně využit pro obchodní část, kdy se lze detailněji zaměřit na často opakovaná témata nebo klíčová slova.

2.5.5. Detekce emocí

Specifickou funkcí, která v poslední době nabývá z pohledu kontaktních center na důležitosti, je možnost rozpoznání emocí vyskytujících se v probíhajícím hovoru. Jednou z variant je zpětná analýza obsahu hovorů z uskutečněných nahrávek nebo v případě dostatečného výpočetního výkonu hovorového systému i online aplikace detekující emoce v hovoru a to v reálném čase.

Detekce emocí probíhá buď na základě vybraných klíčových slov nebo detekcí úrovně hlasitosti řeči v probíhajícím hovoru ve spojitosti s rychlostí mluveného slova. Kombinace těchto metod vede k označení hovorů s potenciálním výskytem emocí, které musí být následně poslechnuty a analyzovány živým člověkem, který může z přesného kontextu hovoru určit závažnost hovoru. Využití analýz emocí vede často ke zlepšení poskytovaných služeb, kdy se ze zpětné analýzy nahrávek hovorů hodnotí a řeší problematické oblasti. Dalším místem, kde tato funkce nachází uplatnění, jsou například služby bezpečnostního nebo ochranného charakteru.

2.6. Nové trendy

S moderními technologiemi je v současnosti možné uskutečnit i dříve futuristické vize, které by nedávnou dobu nikdo neočekával, protože se zdály nerealizovatelné.

Na jedné straně se dosahuje zvýšení produktivity práce kontaktního centra čistě jen nasazením těchto nových funkcí, které umožňují pracovat s hlasem volajícího, a to vhodnou automatizací procesu zpracování přichozího hovoru. Na druhé straně se zákazníkovi dostane lepší služby, kde funkcemi rozpoznání hlasu, se volajícímu dostane relevantnější odpovědi v kratším čase.

Díky dostupnosti těchto nových technologií a trendů je možno v již zmiňovaných call centrech využívat nové funkce založené právě na principech rozpoznání hlasu.

2.6.1. Identifikace hlasem

Hlasová biometrie se v poslední době objevuje čím dál tím častěji jako možnost identifikace volajícího jeho vlastním hlasem a to zvláště v oblastech se zvýšenou bezpečností, jako jsou například banky a pojišťovny nebo instituce, které pracují s osobními daty uživatelů. Současně odpadá nutnost pamatovat si složitá hesla pro komunikaci, uživatel získává rychlejší přístup k požadovaným službám a zároveň se ještě více zvýší bezpečnost přístupu.

Systémy pro svou činnost využívají tzv. zvukový spektrogram, který zohledňuje frekvenci zvuků ve formě vertikální osy a délku ve formě osy horizontální. Kromě toho spektrogram využívá pro znázornění akustické kvality barvy či odstíny šedi. Vlastní systém je pak založen na principu porovnání ověřené vzorové nahrávky a běžného volání, kdy dochází k ověření podobnosti. Díky tomu je možné získat informace anebo provést autorizaci, aniž by musel být konkrétní člověk přítomen. Bezpečnostní systémy k tomu využívají náhodně generované požadavky, které po uživateli vyžadují vyřčení určitých kombinací slov nebo konkrétních vět ještě před začátkem samotného procesu ověřování hlasu. Je důležité si také uvědomit, že funkcionality biometriky řeči (*speech recognition*) není to samé jako hlasová biometrika (*voice recognition*).

2.6.2. Ovládání hlasem

Současná úroveň systémů rozpoznání řeči již na dostatečné úrovni umožňuje detekci samostatných slov, což nabízí možnost mluveným slovem řídit chování aplikací.

V případě volání do call centra formou telefonního hovoru pak tato funkce umožňuje volajícímu směrování hovoru podle jeho přání a to přímo hlasem, a ne tak jako doposud pouze zadáním čísla pomocí tónové volby z telefonu. Kromě jednoduchosti obsluhy se tak významně urychluje zpracování požadavku a následná obsluha. Hlasové systémy jsou schopny automaticky pracovat v několika světových jazycích současně. Samozřejmostí je dnes podpora češtiny včetně specifických slov a dialektů.

2.6.3. Inteligentní směrování příchozích hovorů

Na základě všech informací, které jsou k příchozímu hovoru známé v době jeho zpracování, lze zajistit poměrně inteligentní směrování tohoto hovoru. To se netýká jen konkrétního cíle, kam bude volající přepojen, ale i rychlosti s jakou se dostane k operátorovi kontaktního centra.

Jednou ze základních informací, které k danému hovoru existují, je specifická volba volajícího, kterou provedl na vstupu do kontaktního centra. Jedná se zvláště o výběr služby, kterou chce poskytnout a jeho osobní identifikaci, která urychluje zpracování následného hovoru. Výběr konkrétní služby lze nejčastěji zvolit ve vstupním IVR systému tónovou volbou z telefonu. Identifikaci lze zajistit zadáním osobního čísla a PINu nebo hlasovou identifikací volajícího.

Specifickou technickou informací k danému hovoru je také číslo volajícího porovnatelné s databází existujících kontaktů. Dalším parametrem, který může být využit pro rozhodování o směrování hovoru, je konkrétní čas hovoru nebo den v týdnu (víkend, svátek).

Na základě analýzy všech těchto informací lze rozhodnout, kam a s jakou specifickou prioritou má být konkrétní hovor směrován, zda má být obsloužen konkrétním agentem kontaktního centra (VIP klient) nebo jde o nevhodný hovor, který nemá být v rámci call centra zpracován.

2.6.4. Automatizace odchozích hovorů

V případě aktivního volání zákazníků lze s úspěchem využít zapojení systému detekce hlasu do schématu zpracování hovoru, tak aby se zvýšila efektivita práce této části call centra.

Po vytočení čísla odchozího hovoru, v tuto chvíli ještě bez připojeného agenta kontaktního centra, je hovor analyzován, zda volané číslo není obsazeno nebo případně neexistuje. V případě, že dojde k vyzvánění na volaném čísle, může dojít k přepojení na živého pracovníka nebo je hovor nadále monitorován systémem s připojeným modulem pro rozpoznání hlasu. Ten je schopen, po vyzvednutí do hovorového stavu, eliminovat technická zařízení jako jsou například fax nebo záznamník a pouze v případě hovoru s živou osobou ho dále zpracovat.

V tomto případě se přehraje hláska kontaktního centra s informací o zpracování hovoru a současně dojde k přepojení na jednoho z volných pracovníků, který může hovor zpracovat.

Až do okamžiku vyzvednutí hovoru volanou osobou je tak celý proces hovoru zautomatizován včetně detekce nestandardních stavů. Aplikací této funkce tak dochází k významné úspoře počtu volně dostupných pracovníků call centra resp. zvýšení jejich efektivity.

2.6.5. Převod textu na hlas a řeči na text

V sedmdesátých letech se objevily první technologie, které byly schopny zpracovat hlas a převést ho do formy, která umožnila jeho další analýzu a následné zpracování. Zpočátku se jednalo pouze o jednoduché hlasové ovládání, které bylo významně omezeno dostupnou technickou základnou a výpočetní kapacitou. Vzhledem k technologickému vývoji jsou v současnosti tyto limity překonány a společně s rozvojem výpočetních algoritmů pro zpracování řeči dnes dosahují velmi dobrých výsledků. Důkazem toho je využití těchto funkcí i na úrovni mobilních telefonů, které po převodu hlasu z mikrofону umožní jeho zpracování v interních aplikacích nebo také k ovládání vlastního zařízení.

Obdobně se rozvíjela také oblast převodu textu na hlas, která zpočátku generovala pouze počítačem vytvořený syntetický hlas kvalitou úměrný úrovni v té době dostupné techniky. V průběhu času dospěl vývoj této funkcionality do takového stavu, že v současnosti se již pomocí knihoven hlasu produkuje výsledný zvuk často nerozpoznatelný od hlasu živého člověka. To umožňuje jeho aplikaci na mnoha místech v běžném životě.

Modul převodu řeči na text

S narůstajícím rozvojem systémů pro rozpoznání řeči se blíží doba, kdy kvalita rozpoznání řeči dosáhne takové úrovně, že převod řeči na text a jeho následné zpracování významně změní současně pojetí doposud používaných tradičních systémů.

V současnosti je již podobný systém využíván například pro následnou analýzu telefonních hovorů v případě nahrávek kontaktních centra, kde hlas v těchto hovorech je převeden na text a následně analyzován prostředky pro analýzu textu. V případě této aplikace se účinnost převodu hlasu na text blíží téměř sto procentní úspěšnosti.

Modul převodu textu na řeč

Pro doplnění všech forem automatizovaného zpracování stojí za uvedení již delší dobu využívaný způsob, kdy používaný text je převeden na řeč pomocí předem připravených knihoven hlasu. Ty zabezpečují, že libovolný text je převeden a přehrán jako srozumitelný hlas, který oproti dříve syntetickému hlasu generovaného jen počítačem, je dnes téměř k nerozeznání od reálného lidského hlasu, dokonce ve variantách mužského nebo ženského.

Tuto funkci lze s úspěchem využít například k přečtení textu emailu nebo SMS, pokud není k dispozici nějaké zobrazovací zařízení, případně k předání hlasové informace volajícímu na informační lince například k výši zůstatku na účtu nebo částce platby, případně jiné informace.

3. Praktická část bakalářské práce

V rámci této části práce jsem se zaměřil na zjištění informací, jak je využívání kontaktních center, zvláště jejich telefonní části, vnímáno napříč informační společností a jak ji případně zpětně ovlivňuje, ať v pozitivním nebo negativním směru.

Telefonní služby dnes ke své činnosti využívá většina firem a institucí, a je tak vcelku logickým důsledkem, že se zpracování příchozích nebo odchozích hovorů s veřejností postupně zformovalo do nějaké varianty sofistikované obsluhy těchto hovorů.

Spolu s dostupností moderních technologií jsou tyto služby průběžně doplňovány i o další zařízení, tak aby se obsluha podobných hovorů co nejvíce zautomatizovala, což na jednu stranu vede k možnosti samoobslužného provozu volajícím, na druhé straně vede k významným úsporám na straně kontaktního centra z hlediska menšího množství pracovníků, které takové telefonní hovory obsluhují.

3.1. Výběr metody

Předmětem plánovaného výzkumu bylo získat data z odpovědí na předem připravené otázky, které by po následné analýze měly odpovědět na možné kladené hypotetické otázky. Vzhledem k povaze dat, rozsahu kladených otázek a plánované cílové skupiny se jako optimální jeví průzkum formou kvantitativního výzkumu (Hendl, 2005, s.49), který má následující výhody:

- je vhodný pro zkoumání velkých skupin,
- eliminuje působení rušivých vlivů,
- umožňuje relativně rychlý sběr a analýzu dat,
- výsledek výzkumu je poměrně objektivní, nezávislý na tazateli.

K získání dostatečného množství dat k požadovaným otázkám jsem v rámci pilotního průzkumu zvolil metodu dotazníkového šetření. Tazatel pak samostatně vyplní respektive vybere odpovědi na otázky a to formou interaktivního formuláře umístěného na webovém internetovém portálu. Ten umožňuje obsáhnout dostatečně velký vzorek respondentů tohoto průzkumu napříč celým věkovým spektrem a formou otázek a odpovědí získat zpět informace k následné analýze.

První verze formuláře byla ověřena na menším vzorku tazatelů s cílem otestovat, že formulace otázek a nabízených odpovědí jsou chápány správně. Současně bylo třeba si zkušebně ověřit, že odpovědi z formuláře a jejich rozsah budou dostatečným podkladem pro potvrzení nebo vyvrácení předpokládaných tvrzení ke stanoveným hypotézám a výzkumné otázce.

3.2. Metodika pilotního průzkumu

V metodologické literatuře existují různé definice, popisy a koncepce výzkumu. Tento výzkum lze pojímat jako organizovaný, systematický a logický proces zkoumání, při němž se využívá empirická⁸ informace pro zodpovězení otázek nebo testování hypotéz. Kupříkladu Keith Punch zdůrazňuje ústřední roli výzkumných otázek a systematické používání empirických dat k jejich zodpovězení. Doslova uvádí (Punch, 2015, s.17), že „Výzkum má 4 hlavní rysy:

- zachycuje výzkum pomocí výzkumných otázek nebo testování hypotéz,
- určuje, jaká data jsou k jejich zodpovězení zapotřebí,
- navrhuje výzkum, který by umožnil sběr a analýzu těchto dat,
- využívá data pro zodpovězení výzkumných otázek.“

Mezi mnoha popisy kvantitativního průzkumu je například tento, který je stručný a výstižný:

„Kvantitativní průzkum není nic jiného než testování hypotéz. Teoretický nebo praktický problém je přeložen do jazyka hypotéz, které navrhují, jaké spojení mezi proměnnými bychom měli najít, je-li naše hypotéza pravdivá. Pak následuje sběr dat. Odpovídají-li závislosti mezi sebranými daty vzorci předpovězenému v hypotéze, přijmeme je jako platné. Jinak je musíme odmítnout.“

(Disman, 2007, s.76)

Dále také v průběhu výzkumu uvádí, že: *„Opomenutí pilotní studie nás v některých případech nemusí vytrést vážným zkršením, zatímco opomenutí předvýzkumu je nesmírně riskantní. Předvýzkum⁹ by měl být součástí každé výzkumné akce.“*

(Disman, 2007, s.122)

3.3. Volba průzkumného vzorku

Cílem plánovaného průzkumu je získat přehlednou informaci v co nejširším věkovém spektru a v jeho rámci se k požadovanému tématu dotázat na konkrétní otázky. Aby měl průzkumný vzorek správnou vypovídací hodnotu, je optimální docílit v jednotlivých věkových kategoriích stejného poměru respondentů tak jako je poměrně zastoupen v rámci celé České republiky.

Žádost o vyplnění interaktivního dotazníku byla rozeslána všem věkovým skupinám v poměrech přibližně odpovídající běžnému zastoupení v populaci. V případě, že návratnost z dané věkové kategorie nebyla dostatečná, oslovení respondentů z této věkové skupiny se opakovalo. Celkově

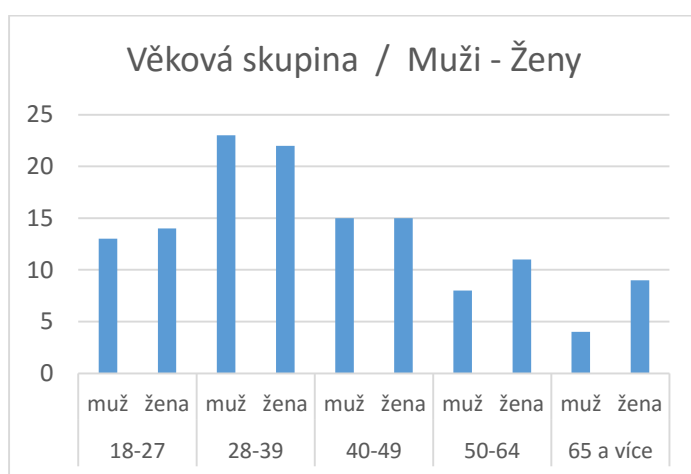
⁸ Empirismus je filozofický pojem, který považuje zkušenost za základ nebo zdroj znalostí

⁹ Předvýzkum – ověřuje vhodnost a adekvátnost navrhovaného výzkumného postupu jako celku

se tohoto pilotního průzkumu zúčastnilo 134 respondentů, s požadovaným množstvím zastoupených mužů a žen, od kterých jsem v dotazníku obdržel kompletní data.

Jak vyplývá z informací Českého statistického úřadu ¹⁰ pro rok 2015, je poměr počtu mužů a žen v běžné populaci vyrovnaný, resp. mladších mužů je oproti ženám více v poměru 100:95, ale naopak přibližně od věkové hranice 65 let počet mužů klesá až do poměru 2:1 ve prospěch žen. Je tedy vhodné, aby i v průzkumném vzorku odpovídalo množství mužů a žen v daných věkových kategoriích jejich poměrnému zastoupení v reálné populaci.

Následný graf zobrazuje poměr zastoupení jednotlivých věkových skupin celého průzkumného vzorku a tabulka pak jejich absolutní počet:



Věková skupina	Počet
18-27	27
muž	13
žena	14
28-39	45
muž	23
žena	22
40-49	30
muž	15
žena	15
50-64	19
muž	8
žena	11
65 a více	13
muž	4
žena	9
Celkový součet	134

Obrázek 9: Zastoupení respondentů v průzkumném vzorku

Jak je z počtů v tabulce patrné, podařilo se pro jednotlivé věkové kategorie docílit vhodného poměru žen a mužů s důrazem na vyšší počet žen oproti mužům ve vyšším věku.

Jak již bylo uvedeno, cílem tohoto průzkumného vzorku bylo sestavit množství respondentů v jednotlivých věkových kategoriích tak, aby konaný pilotní projekt a jeho výsledná data co nejlépe odpovídala věkovému zastoupení všech věkových skupin. K tomuto účelu jsem dohledal údaje dostupné z Českého statistického úřadu a to jak přesné množství obyvatel žijících na území ČR, tak jejich zastoupení v daných věkových kategoriích.

Základní informaci o respondentovi jako je jeho věk a pohlaví je také možno doplnit o úroveň dosaženého vzdělání, a tyto informace následně porovnávat s dalšími odpověďmi daného průzkumného šetření.

¹⁰ https://www.czso.cz/csu/xp/analyza-_zeny_a_muži_v_krajich_cr_demografie

Zde jsou uvedeny reálné číselné údaje (data ČSÚ k 31. 12. 2014) :

Věk	Počet obyvatel na území ČR	Poměr skupiny k počtu obyvatel	Optimální počet ve vzorku	Reálný počet ve vzorku
18-27	1.275.172	12,1%	24,6	27
28-39	2.144.583	20,4%	41,3	45
40-49	1.486.743	14,1%	28,6	30
50-64	2.065.260	19,6%	39,8	19
				13 (nad 65 let)
Celkem	10.538.275	100%	134,3	134

Uvedená tabulka zobrazuje jak absolutní počty obyvatel v jednotlivých věkových kategoriích, tak jejich poměr k celkovému počtu obyvatel na území ČR. Tento poměr je pak vztažen k počtu 134 osob, které se zúčastnily pilotního průzkumu, a zobrazuje tak jejich optimální počet. V posledním sloupci je pak reálný počet těchto osob, tak jak svůj věk v průzkumu uvedli.

Ve věkových kategoriích od 18 do 49 let se podařilo poměrně přesně přiblížit k požadovanému stavu a reálný počet respondentů tak odpovídá optimálnímu poměru mužů a žen. V kategoriích nad 50 let se shromáždit data od dostatečného počtu respondentů nepodařilo, tak aby přesněji odpovídal požadované hodnotě zastoupení v dané věkové kategorii. V optimálním stavu by měl být počet respondentů přibližně dvojnásobný. Nicméně pokud bychom pro tento pilotní průzkum započítali data od skupiny nad 65 let i jako respondenty skupiny věku 50-64 let, pak by tento vzorek reprezentoval data již od 32 lidí, což se již blíží optimálnímu požadovanému stavu v počtu 39 respondentů. V reálném průzkumu bychom se pak zaměřili na cílené oslovení této věkové skupiny, tak abychom získali odpovědi od požadovaného počtu respondentů.

Jako respondenti byly při uskutečněném průzkumu osloveny osoby jak z mého nejbližšího okolí, tak spolupracovníci, spolužáci ze současného i předchozího studia a jejich rodinní příslušníci. Tím bylo dosaženo oslovení přibližně stejného množství mužů a žen v každé věkové kategorii.

Cílem průzkumu nebyla žádná specifická ani odborná skupina, ale běžná osoba se zkušeností s voláním do nějakého kontaktního centra, kterou je občas z nějakého důvodu téměř každý z nás. Hlavním cílem bylo zjistit, jak informace k stávajícímu chování uživatelů, tak jejich možným požadavkům k funkcím zaměřených na hlas.

3.4. Definice výzkumné otázky a hypotézy

V rámci výzkumu je třeba formulovat hlavní výzkumnou otázku, která vymezuje to, na co chceme znát odpověď v souvislosti s řešeným problémem. V případě této práce na téma hlasových systémů je možné stanovit následující otázku:

Jaký je vztah mezi věkem a ochotou používat funkce s analýzou hlasu ?

Důležitou částí je pak stanovení hypotéz, ke kterým hledáme odpovědi. Stručně můžeme hypotézu popsat jako domněnku odvozenou z teorie nebo zkušenosti, kterou lze empiricky zpochybnit. Vlastní pojem slova hypotéza je poměrně přesně definován a například Hendl ji popisuje takto:

„Hypotéza je tvrzení o podstatě určité situace ve zkoumané oblasti. Jde o návrh nebo představu o stavu mezi uvažovanými výzkumnými proměnnými. Tento návrh se výzkumem zamítá nebo potvrzuje a to na základě empirické evidence (faktů, dokladů). Hypotézy mohou být navrženy na základě předběžného výzkumu, tedy pomocí pilotní studie, jež se provádí za účelem zmapování problému a získání předběžných informací. Můžeme pak formulovat několik hypotéz pro danou výzkumnou otázku.“ (Hendl, 2015, s.25)

Pro průzkum v této části práce si tak definuji následující hypotézy:

H1: Je uživatelům známo, že hlas lze využít k osobní identifikaci?

H2: Chtějí mladší uživatelé více používat ovládání hlasem?

H3: Byli by uživatelé ochotni využívat identifikaci hlasem jako placenou službu?

Pro testování hypotézy se využívá postup statistického testování, který využívá dvou pojmů:

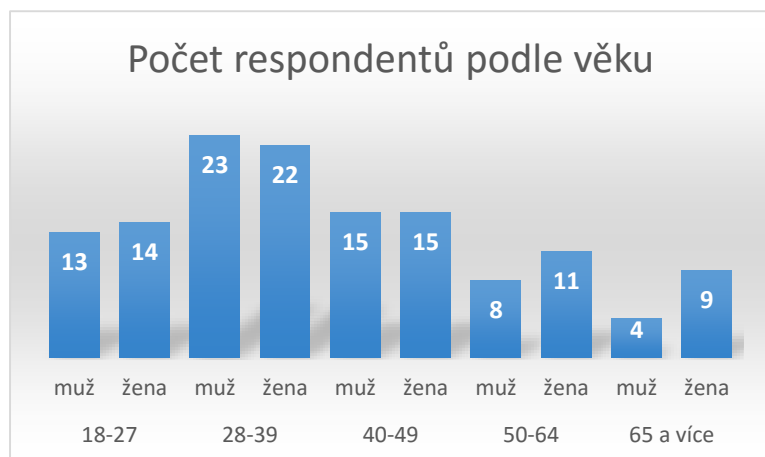
- *Nulová hypotéza, značená obvykle jako H_0 , je tvrzení které deklaruje „téměř žádný rozdíl“. Je to jednoduchá myšlenka, že se ve zkoumaném vzorku nic neděje a neexistují tu žádné rozdíly, že neexistuje žádný vztah mezi hodnotami, které jsme naměřili. To je hypotéza, kterou by výzkumník rád spíše zamítl a lze vymezit způsobem, že rozdíl nedosahuje určité hodnoty.*
- *Alternativní hypotéza, označovaná jako H_1 , znamená situaci kdy nulová hypotéza H_0 neplatí. Je to představa, že „se něco děje“, a obvykle se vyjadřuje jako existence difference.*

Pokud jsme si definovali nulovou hypotézu ve tvaru, že rozdíl nedosahuje požadované hodnoty, pak uvažujeme alternativní hypotézu jako platnou. Pokud nedokážeme opak, pak budeme předpokládat, že platí nulová hypotéza. (Walker, 2013)

3.5. Prezentace otázek a odpovědí

Věkové zastoupení

Pilotního průzkumu se zúčastnilo celkem 134 respondentů a následující graf přehledně zobrazuje početní zastoupení v jednotlivých věkových kategoriích:



Obrázek 10:

Počty respondentů podle věku

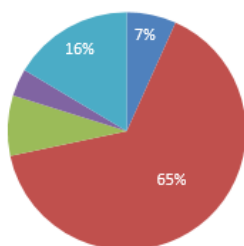
Uvedený počet přibližně odpovídá demografickému zastoupení mužů a žen v běžné populaci a může tak reprezentovat informativní vzorek pro dané téma.

Využívání smart fonu

Jednou z důležitých otázek byl dotaz, zda respondent využívá chytrý telefon (smart phone), případně jestli zná funkci převodu řeči na text a k čemu ji v telefonu používá.

Nějakou formu chytrého telefonu používá 84% uživatelů a překvapivě více než 3/4 ze všech uživatelů ví, že tento typ telefonu obsahuje funkci na zpracování řeči. Aktivně ji však v telefonu, například k psaní zpráv nebo ovládání tohoto typu zařízení, využívá jen menší část, cca 12 %. Ostatní uživatelé ji k žádné specifické funkci nepoužívají.

Používáte chytrý telefon (smart phone)? Víte o funkci převodu řeči na text?

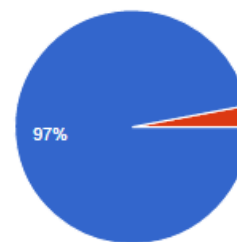


Ano, ale o této funkci nevím	9	7%
Ano, o této funkci vím, ale nepoužívám	87	65%
Ano, tuto funkci používám pro psaní zpráv	11	8%
Ano, tuto funkci používám v maximální míře (zprávy, email, ovládání telefonu)	5	4%
Ne, smart phone nepoužívám	22	16%

Používání emailu

Další doplňující otázkou, s cílem obdržet komplexní informaci, byl dotaz: **Používáte email ?**

Není bez zajímavosti, že 4 respondenti z celkového počtu dotázaných uvedli, že žádný email nepoužívají. To představuje přibližně 3% ze všech respondentů. K vyplnění dotazníku, který byl koncipován formou online formuláře na webové stránce, nebylo použití emailu nijak zapotřebí. Vše se realizovalo pouze pomocí webového prohlížeče a to online.



V době, kdy jeden člověk používá i více emailů, je tato informace přinejmenším zajímavá.

Dosažené vzdělání

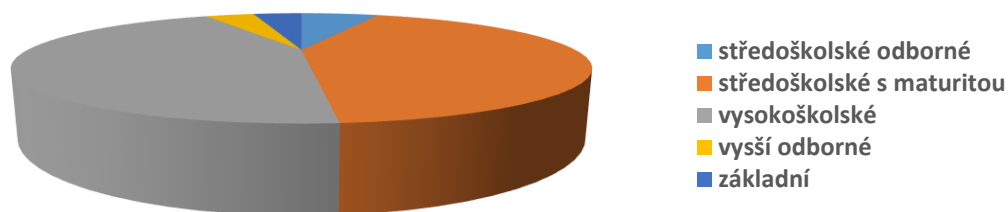
Jednou ze základních informací o respondentovi je také úroveň dosaženého vzdělání, kterou si každý v dotazníku mohl zvolit z nabízené škály a to od vysokoškolského až po základní.

Největší část respondentů uvedla vysokoškolské vzdělání, částečně bylo zastoupeno i vzdělání vyšší

odborné, druhou největší skupinou pak reprezentují středoškoláci s maturitou. Všichni ti, kteří uvedli základní vzdělání, zároveň uvedli svůj věk odpovídající nejmladší věkové skupině, takže se s největší pravděpodobností jedná o ještě studující středoškoláky.

Nejvyšší dosažené vzdělání	Počet
vysokoškolské	59
vyšší odborné	5
středoškolské s maturitou	57
středoškolské odborné	8
základní	5
Celkový součet	134

Skupiny respondentů podle vzdělání

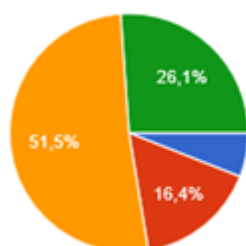


3.5.1. Využívání služeb kontaktních center

Četnost volání

V tomto pilotním průzkumu byl jednou ze základní otázek, zda lidé využívají služeb některého z kontaktních center a pokud ano, jak často v průběhu roku se tak děje.

Využíváte ke komunikaci s institucemi jejich kontaktních center? Jak často?



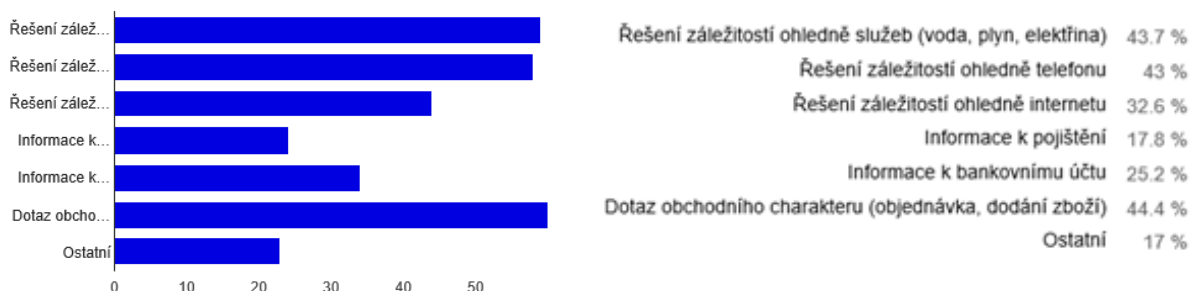
Četnost volání	Číslo respondentů	Podíl (%)
vícerať v měsíci	8	5.9 %
1x měsíčně	22	16.3 %
vícerať za rok	69	51.5 %
1x za rok	35	26.1 %

Pouze jedenkrát za rok volá do nějakého kontaktního centra čtvrtina respondentů. Ostatní dotazovaní volají častěji v průběhu roku, přibližně čtvrtina z nich nejméně jednou měsíčně.

Rozsah požadovaných služeb

Další z důležitých informací k uskutečněným hovorům do kontaktního centra je důvod nebo konkrétní potřeba volajícího řešit nějakou konkrétní záležitost. V následujícím grafu jsou zachyceny počty těchto požadavků a jejich obsah:

Kvůli čemu voláte nejčastěji?

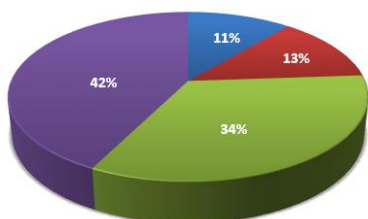


Z tohoto pilotního průzkumu vyplynulo, že nejčastější důvod volání na linku kontaktního centra má obchodní charakter, kde se jedná například o objednání zboží nebo jeho dodání. Stejně častým důvodem je řešení záležitostí spojených se službami k telefonu nebo dodávek energií (voda, plyn, elektro). Přibližně poloviční četnost pak má volání do bank, které dříve voláním do kontaktních center dominovalo. V současnosti je ale významně nahrazeno internetovým bankovníctvím, takže množství telefonních hovorů se oproti minulosti snížilo.

3.5.2. Způsob používání

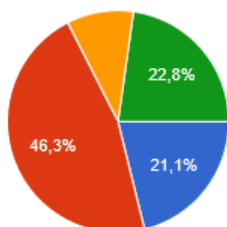
Jedna z částí pilotního průzkumu se soustředila na zjištění preferencí uživatelů, jak k volbě možných služeb kontaktního centra, tak urychlení obsluhy využitím automatizovaných systémů. Konkrétně se jednalo o následující otázky s těmito výsledky:

Jak reagujete, pokud informační linka požaduje volbu aktivity pomocí čísel?



Nedělám nic a vyčkám přihlášení pracovníka kontaktního centra	15	11%
Používám jen první volbu a pak čekám	17	13%
Využiji více voleb, avšak max 3-4, pak už čekám	45	34%
Využívám maximálně s cílem co nejdříve získat informaci	57	42%

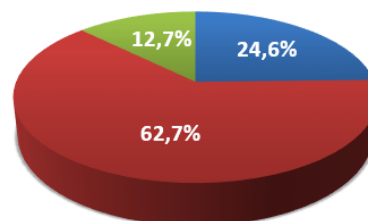
Pokud je možno zadat clientské číslo, využíváte ho ke své identifikaci?



Vždy využívám	31	21.1 %
Jen, když to urychluje moje odbavení	62	46.3 %
Jen, pokud mám zadat jedno číslo	12	9.8 %
Nezadávám	28	22.8 %

Jaký typ komunikace s kontaktním centrem preferujete?

Vždy chci komunikovat s živým člověkem, i když budu čekat déle	84	62,7%
Preferuji obsluhu volbou čísel, pokud je to rychlejší než čekat	33	24,6%
Vždy preferuji obsluhu volbou čísel, hovor chci jen při problému	17	12,7%



Z uvedených reakcí respondentů vyplývá, že pokud jejich aktivita může urychlit získání informace nebo zkrátit čas čekání, jsou ochotni spíše být aktivní než čekat a nedělat nic.

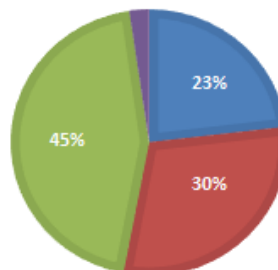
Například téměř 40% respondentů využívá maximálně volbu pomocí čísel s cílem co nejdříve získat požadovanou informaci. Pokud systém umožňuje zadání clientského čísla, pak téměř polovina uživatelů využije tuto možnost s cílem urychlit svoje odbavení.

Nicméně téměř 2/3 všech respondentů stále preferují komunikaci s živým pracovníkem kontaktního centra i přesto, že budou pravděpodobně déle čekat.

Některá kontaktní centra nabízejí, a to v případě delšího čekání ve frontě hovorů, možnost zvolit funkci zpětného volání, která místo dalšího čekání předá tento požadavek a hovor se ukončí. V nejbližší možnou dobu je pak uživateli zavoláno zpět na jeho původní číslo a tak se eliminuje čas jeho čekání ve frontě, čímž se docílí zvýšení spokojenosti uživatelů daného call centra.

Cílem otázky v dotazníku bylo zjistit informace k této funkci:

Ne, vždy čekám na živého pracovníka	31	23%
O této možnosti nevím	40	30%
Využívám jen, pokud je čekací čas příliš dlouhý	60	45%
Využívám vždy	3	2%



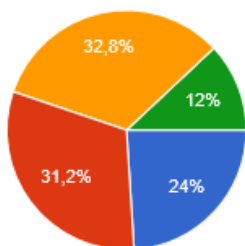
Přibližně třetina respondentů se s touto funkcí pravděpodobně nesešla. Další čtvrtina by ji ani využít nechtěla, protože preferuje hovor s živým pracovníkem i za cenu delšího čekání. Nicméně téměř polovina ji už využívá a to zvláště v případě, že je čekací čas příliš dlouhý.

3.5.3. Volba funkčnosti hlasem

V současnosti se stále častěji objevuje možnost ovládat chování infolinky přímo vlastním hlasem namísto ovládání tónovou volbou pomocí stisku číslic na telefonu.

Výsledek dotazu v pilotním průzkumu je následující:

Přivítali byste možnost volit chování infolinky pomocí hlasu?



Ano	30	24 %
Spíše Ano	39	31.2 %
Spíše Ne	41	32.8 %
Ne	15	12 %

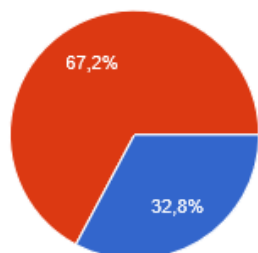
Více než polovina respondentů, a to téměř 55%, by tuto možnost uvítala. O podobnou funkci nemá zájem menší část respondentů, ale odpověď „Ne“ dalo v tomto případě pouze 12% lidí.

Funkce rozpoznání hlasu zažívá v poslední době renezanci svého používání. Dříve byla limitována technickými prostředky a vývojem software, což se v průběhu času významně změnilo ve prospěch funkčnosti této technologie. Budoucnost teprve ukáže její reálné využití.

3.5.4. Identifikace hlasem

Hlas každého člověka je jedinečný a lze ho tak využít k osobní identifikaci. V případě jeho využití v prostředí kontaktních center by tak mohl přímo nahradit v současnosti používaná osobní identifikační čísla. Ta se standardně zadávají pomocí tónové volby stiskem tlačítek telefonu. Ke znalosti této funkce byl vznesen tento dotaz:

Víte o možnosti, že váš hlas lze využít k osobní identifikaci?

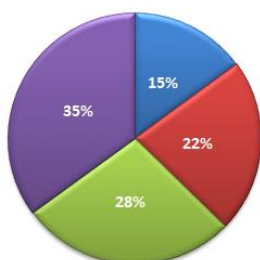


Ano, vím	47	32.8 %
Ne, nevím	87	67.2 %

Je zajímavé, že pouze 1/3 respondentů o této funkci ví nebo o ní alespoň slyšela.

Následující otázka se dotazuje, zda by respondent chtěl tuto funkci využívat, pokud by zjednodušila jeho odbavení v rámci obsluhy infolinky kontaktního centra:

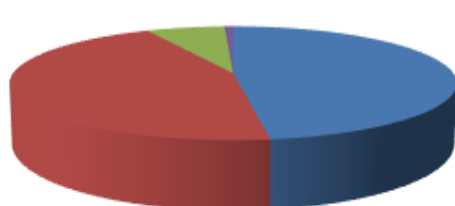
Pokud identifikace hlasem zjednoduší vaše odbavení, preferovali byste tu možnost?



Ano	30	22%
Spíše Ano	47	35%
Spíše Ne	37	28%
Ne	20	15%

Větší polovina dotázaných, přesněji 57% všech respondentů, by tak možnost identifikace pomocí hlasu preferovala a zjednodušila by si tak své odbavení.

Na otázku, zda by byli **ochotni platit za službu, pokud by se urychlilo získání informace** nebo umožnilo přednostní odbavení, a to díky identifikaci hlasem respondenti většinou odpověděli negativně a to v polovině odpovědí „Ne“ a „Spíše Ne“ v dalších 44 procentech.



Ne	64	47,8%
Spíše Ne	60	44,8%
Spíše Ano	9	6,7%
Ano	1	0,7%

3.6. Vyhodnocení odpovědí

Provedený pilotní průzkum poskytl dostatečné množství informací, aby bylo možno odpovědět na otázky definované před jeho započítím. V této části tedy s ohledem na obdržená data určím, zda se předpokládané hypotézy potvrdily či nikoli.

U otázky **“Zda je uživatelům známo, že hlas lze využít k osobní identifikaci”** odpověděly 2/3 dotázaných, že o této možnosti **NEVĚDÍ**.

Analýzou dat v kombinaci s věkem bylo zjištěno, že informovanost o této funkci klesá s přibývajícím věkem. U věkové skupiny od 18 do 27 let totiž o této funkci věděla téměř polovina zúčastněných, zatímco v kategorii nad 65 let už to bylo jen čtvrtina dotázaných respondentů.

Mladší věkové kategorie jsou tak o této funkci informovány lépe.

Věková kategorie	Počet
18-27	27
Ano, vím	12
Ne, nevím	15
28-39	45
Ano, vím	15
Ne, nevím	30
40-49	30
Ano, vím	11
Ne, nevím	19
50-64	19
Ano, vím	6
Ne, nevím	13
65 a více	13
Ano, vím	3
Ne, nevím	10
Celkový součet	134

Další otázkou byla hypotéza vázaná na věk respondenta:

Chtějí mladší uživatelé více používat ovládání hlasem?

Počet dotazovaných, kteří na otázku k preferenci ovládání infolinky hlasem odpověděli kladně a to „Ano“ nebo „Spíše ano“ byl 76 z celkového počtu 134 všem respondentů.

Vyjádřeno v procentech se kladně vyjádřilo 57% dotázaných.

Ve věkové kategorii 18-27 let pak činila tato hodnota 74%,

t.j. téměř 3/4 mladších uživatelů by ovládání hlasem

preferovalo. Odpověď na tuto otázku je tedy **ANO**.

Věková kategorie	Počet / Celkem	Poměr
18-27	20 / 27	74%
Ano	8	
Spíše Ano	12	
28-39	25 / 45	56%
Ano	10	
Spíše Ano	15	
40-49	18 / 30	60%
Ano	8	
Spíše Ano	10	
50-64	7 / 19	37%
Ano	4	
Spíše Ano	3	
65 a více	6 / 13	46%
Ano	3	
Spíše Ano	3	
Celkový součet	76 / 134	57%

Třetí hypotézou, na kterou měl pilotní průzkum najít odpověď, byla následující otázka:

Byli by uživatelé ochotni využívat identifikaci hlasem jako placenou službu?

Odpověď na tuto otázku byla celkem jednoznačná a to, že **NE**.

Pouze 10 respondentů by uvažovalo, že by za tuto službu bylo ochotno zaplatit. Větší část respondentů tohoto průzkumu o tuto službu zájem nemá. Odpověď „Ne a „Spíše Ne“ zvolilo více než 92% respondentů z celkové množství.

Odpověď	Počet	Procenta
Ne	64	47,8%
Spíše Ne	60	44,8%
Spíše Ano	9	6,7%
Ano	1	0,7%
Celkový počet	134	100,0%

Poslední otázkou, na kterou hledal průzkum odpovědi, byla tato formulace:

Jaký je vztah mezi věkem a ochotou používat funkce s analýzou hlasu ?

Na základě předchozích odpovědí na hypotézy a současně komplexních informací z pilotního průzkumu k tématu kontaktních center lze konstatovat, že vazba mezi věkem a větší ochotou používat funkce s analýzou hlasu je viditelná spíše u mladší věkové kategorie. Téměř polovina mladších respondentů věděla o možnosti identifikovat se hlasem, což je téměř dvojnásobek oproti starším uživatelům kontaktních center. Současně téměř 3/4 těchto mladších uživatelů by ovládnutí hlasem preferovalo jako možnost chování infolinky oproti stávajícímu stavu.

Mladší uživatelé kontaktních center jsou ochotnější k používání funkcí spojených s analýzou hlasu. To se týká jako identifikace hlasem, tak možností hlasem ovládat chování aplikací, které řídí chování kontaktního centra a tudíž i nabídky jeho služeb.

Toto konstatování možná jen potvrzuje fakt, že mladší lidé ve větší míře než ostatní používají smart phony, které v dnešní době již funkce spojené s používáním hlasu obsahují. Nemají tedy z této technologie obavu, ale naopak ji aktivněji vyhledávají.

4. Závěr

Obsahem mojí bakalářské práce bylo popsat téma funkcí s rozpoznáním řeči a to jak teoreticky, tak i v rámci praktické části. Těmito funkcemi jsou myšleny zejména identifikace hlasem a rozpoznání řeči s možným převodem řeči na text, se kterým lze dále pracovat.

Zvláště v praktické části této práce byl kladen důraz na zjištění, jestli se volající do kontaktních center setkávají s obtížnější obsluhou, která jim ztěžuje nebo prodlužuje získání požadované informace a zda by využití funkcí pro rozpoznání řeči tento problém pomohlo odstranit.

Z empirické části průzkumného šetření vyplynulo, že zvláště mladší věkové kategorie jsou uživateli, kteří by tyto technologie preferovali. Pro starší uživatele je stávající obsluha někdy obtížnější, přesto ale funkce spojené s rozpoznáním hlasu nepreferují, a to i za cenu delšího čekání, kdy používají jim aktuálně známý systém bez požadovaných změn. Z hlediska služby pak často preferují jen obsluhu živým pracovníkem kontaktního centra oproti vlastní aktivitě.

Využívání technologií pro rozpoznání hlasu tak jednoznačně preferuje spíše mladší věková kategorie a to do věku 28 let. Pokud je tedy zájem významněji tyto technologie uplatnit, uživatelé tohoto věku jsou pak tou správnou cílovou skupinou, která tento přístup ocení.

Budoucí rozvoj technologií, spojených s rozpoznáním hlasu, bude směřovat do všech věkových kategorií, tak aby jeho použití bylo na věku nezávislé. Jen tak se uplatní napříč celou informační společností, ve které v současnosti žijeme. Cílem toho trendu je všudypřítomnost hlasového ovládání, kde klávesnici nahradí hlas a jakékoli zařízení na něj bude schopno plně reagovat. Rozpoznávání řeči tak bude uživatelům k dispozici kdykoli a kdekoli.

5. Seznam použité literatury

BCC Research [online]. Voice Recognition, Technologies and Global Markets. BCC Research LLC, 2010 [cit. 2016-04-29]. Dostupné online z:

<http://www.bccresearch.com/report/voice-recognition-tech-markets-ift039b.html>

ČESKÝ STATISTICKÝ ÚŘAD [online]. *Věkové složení obyvatel k 31.12.2014*. ČR, Praha.

Dostupné online z: <https://www.czso.cz/csu/czso/vekove-slozeni-obyvatelstva-2014>

DISMAN, Miroslav. *Jak se vyrábí sociologická znalost: příručka pro uživatele*.

Praha: Karolinum, 2007. ISBN 80-246-0139-7.

HENDL, Jan. *Kvalitativní výzkum: základní metody a aplikace*. Praha: Portál, 2005, 407 s.

ISBN 80-7367-040-2.

HENDL, Jan. *Přehled statistických metod: analýza a metaanalýza dat*. Páté, rozšířené vydání.

Praha: Portál, 2015. ISBN 978-80-262-0981-2.

KOCIÁNOVÁ, Pavla. *Využití simulace k analýze chodu modré linky Karneval*.

Praha: Vysoká škola ekonomická, 2007. 62s. Diplomová práce. Dostupný také online z:

<http://isis.vse.cz/lide/clovek.pl?zalozka=13;id=21862;studium=60043;zp=5904>

LEITNER, Jakub. *Hlasové kodéry pro nízké přenosové rychlosti*. Brno: VÚT, 2009. 72s.

Diplomová práce. Dostupný také online z:

https://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=14238

NOUZA, Jan. Využití hlasových technologií v praxi. In: *Uživatelsky přívětivá rozhraní*.

Praha: Horava & Associates, 2009, 177 s. ISBN 978-80-254-5295-0

PAPÍK, Richard. *Vyhledávání informací II. Uživatelské rozhraní a vlivy oboru „Human-Computer Interaction“*. Národní knihovna: knihovnická revue [online], 2001 [cit. 2016-04-29].

Dostupný také online z: <http://full.nkp.cz/nkkr/NKKR0102/0102081.html>

PAPÍK, Richard. Bibliometrické a infrometrické metody v databázích. In: *Analýza a zpracování signálů V : semináře katedry teorie obvodů*. Praha: ČVUT, 2004, s. 168-175.

ISBN 80-01-03139-X.

PAPÍK, Richard, Martin SOUČEK, Barbora DROBÍKOVÁ. Problematika HCI v rešeršních systémech. In: *Uživatelsky přívětivá rozhraní*. Praha: Horava & Associates, 2009, 177 s. ISBN 978-80-254-5295-0

PUNCH, Keith. *Úspěšný návrh výzkumu*. Vyd. 2. Praha: Portál, 2015. 230 s. Z anglického originálu *Developing effective research proposals* přeložil Jan Hendl. ISBN 978-80-262-0980-5

SKARNITZL, Radek, ŠTURM, Pavel a VOLÍN, Jan. *Zvuková báze řečové komunikace: fonetický a fonologický popis řeči*. První vydání. Praha: Univerzita Karlova v Praze, Nakladatelství Karolinum, 2016. 170 stran. ISBN 978-80-246-3272-8

SMOLOVÁ, Iva. *Fluktuace zaměstnanců call centra*. Praha: Univerzita Karlova v Praze, 2011. 107 s. Diplomová práce. Dostupný také online z: http://ckis.cuni.cz/F/?func=direct&doc_number=001289022

SPRIGG, Christine; SMITH, Phoebe; JACKSON, Paul. *Psychosocial risk factors in call centres: an evaluation of work design* [online]. Sudbury: HSE, 2003 [cit. 2016-04-29]. Dostupné z: <http://www.hse.gov.uk/research/rrpdf/rr169.pdf> . ISBN 0-7176-2774-8.

THE HISTORY OF THE CALL CENTRE. In: *Call centre helper* [online]. Call Centre Helper, 2011 [cit. 2016-04-29]. ISSN 1718-892X. Online dostupné z: <http://www.callcentrehelper.com/the-history-of-the-call-centre-15085.htm>

WALKER, Ian. *Výzkumné metody a statistika*. Vyd. 1. Praha: Grada, 2013. 218 s. Z pohledu psychologie. ISBN 978-80-247-3920-5.

YEMCHENKO, Viktoriia. *Moderní technologie pro ovládání PC hlasem*. Praha, VŠE, 2011. Dostupný také online z: <http://info.sks.cz/www/zavprace/soubory/72810.pdf>

6. Seznam obrázků

Obrázek 1	Uživatelské rozhraní Proquest	129
Obrázek 2	Zobrazení vyhledaných prací v čase (Proquest, 2016).....	129
Obrázek 3	Uživatelské rozhraní IEEE Xplore	13
Obrázek 4	Zobrazení vyhledaných prací v čase (IEEE, 2016).....	13
Obrázek 5	Uživatelské rozhraní Web Of Science.....	143
Obrázek 6	Zobrazení citací vyhledaných prací.....	143
Obrázek 7	Trh s aplikacemi pro rozpoznání řeči (BCC Research, 2010).....	43
Obrázek 8	Etapy rozpoznávání řeči	18
Obrázek 9	Zastoupení respondentů v průzkumném vzorku.....	45
Obrázek 10	Počty respondentů podle věku.....	46

Příloha I.

Online dotazník

Níže je uvedena kopie online dotazníku, tak jak byl k dispozici v rámci dotazníkového šetření, a to se všemi otázkami a nabízenými odpověďmi:

Využíváte kontaktní centra ?

Jste? *

- žena
 muž

Vaše dosažené vzdělání *

▼

- základní
- středoškolské odborné
- středoškolské s maturitou
- vysší odborné
- vysokoškolské

Váš věk *

▼

- Méně než 18 let
- 18-27
- 28-39
- 40-49
- 50-64
- 65 a více

Používáte email? *

- Ano
 Ne

Používáte chytrý telefon (smart phone)? Víte o funkci převodu řeči na text? *

(Týká se platforem Android, iOS, Windows Mobile)

- Ne, smart phone nepoužívám
 Ano, ale o této funkci nevím
 Ano, o této funkci vím, ale nepoužívám
 Ano, tuto funkci používám pro psaní zpráv
 Ano, tuto funkci používám v maximální míře (zprávy, email, ovládání telefonu)

Využíváte ke komunikaci s institucemi jejich kontaktních center? Jak často? *

- vícekrát v měsíci
- 1x měsíčně
- vícekrát za rok
- 1x za rok

Kvůli čemu voláte nejčastěji? *

- Řešení záležitostí ohledně služeb (voda, plyn, elektřina)
- Řešení záležitostí ohledně telefonu
- Řešení záležitostí ohledně internetu
- Informace k pojištění
- Informace k bankovnímu účtu
- Dotaz obchodního charakteru (objednávka, dodání zboží)
- Other:

Jaký typ komunikace s kontaktním centrem preferujete? *

- Vždy chci komunikovat s živým člověkem, i když budu čekat déle
- Preferuji obsluhu volbou čísel, pokud je to rychlejší než čekat
- Vždy preferuji pouze obsluhu volbou čísel, osobní hovor chci jen v případě problému

Jak reagujete, pokud informační linka požaduje volbu aktivity pomocí čísel? *

- Nedělám nic a vyčkám přihlášení pracovníka kontaktního centra
- Používám jen první volbu a pak čekám
- Využiji více voleb, avšak max 3-4, pak už čekám
- Využívám maximálně s cílem co nejdříve získat informaci

Pokud je možno zadat klientské číslo, využíváte ho ke své identifikaci? *

- Vždy využívám
- Jen, když to urychluje moje odbavení
- Jen, pokud mám zadat jedno číslo
- Nezadávám

Pokud čekáte, využíváte možnost místo čekání být zván zpět? *

(Některá kontaktní centra nabízejí tuto funkci jako volbu v čekací frontě)

- Ne, vždy čekám na živého pracovníka
- Využívám jen, pokud je čekací čas příliš dlouhý
- Využívám vždy
- O této možnosti nevím

Přivítali byste možnost volit chování infolinky pomocí hlasu? *

(Namísto ovládání číselnou volbou)

- Ano
- Spíše Ano
- Spíše Ne
- Ne

Víte o možnosti, že váš hlas lze využít k osobní identifikaci? *

(Místo identifikace klientským číslem a heslem)

- Ano, vím
- Ne, nevím

Pokud identifikace hlasem zjednoduší vaše odbavení, preferovali byste tu možnost? *

(Místo identifikace klientským číslem a heslem)

- Ano
- Spíše Ano
- Spíše Ne
- Ne

Pokud by vaše identifikace hlasem urychlila získání informace a umožnila přednostní odbavení, byli byste ochotni za ni platit jako za službu? *

- Ano
- Spíše Ano
- Spíše Ne
- Ne

Odeslat

Po stisku tlačítka „Odeslat“ se dotazovanému objevila potvrzující zpráva o dokončení současně s možností se pro informaci podívat na souhrn již dokončených dotazníků.

Vaše odpovědi byly zaznamenány. Děkujeme

[Zobrazit souhrn odpovědí](#)

[Odeslat další odpověď](#)