CONTRIBUIÇÕES PARA *CLOSED CAPTION* AUTOMÁTICO APLICADO EM SISTEMAS DE TV DIGITAL

UNIVERSIDADE ESTADUAL DE CAMPINAS – FACULDADE DE TECNOLOGIA

Clayton Francisco Santana Junior (Bolsista) e Prof. Dr. Rangel Arthur (Orientador),
Faculdade de Tecnologia - FT, UNICAMP. Agência Financiadora: CNPq

Palavras-Chave: Tv digital - Closed Caption - Acessibilidade

Introdução

De acordo com o censo demográfico 2010 IBGE, cerca de 5,1% da população brasileira possui deficiência auditiva. Para tal integração dessas pessoas na mídia televisiva brasileira é necessária a utilização de legenda oculta (*closed caption*), como recurso de acessibilidade para o público em situações específicas como: ambientes que necessitam de silêncio, e ambientes barulhentos. Para viabilizar tal recurso durante a ocorrência de fala espontânea ao vivo e gravada, é apresentada a tecnologia de reconhecimento automático de voz incorporada junto aos equipamentos das emissoras de TV.

Neste trabalho, foi comparado diferentes softwares disponíveis de forma gratuita para a conversão de sinal de áudio em texto. Diante dos mesmos, foram investigados diferentes tipos de sinais de vídeo digital e encapsulamento de dados de *closed caption* no sinal digital.

Metodologia

O desenvolvimento para implantar o sistema é abordado onde, o sinal captado é duplicado e inserido em um servidor com o software de reconhecimento por voz. A seguir o texto gerado e a imagem captada pela câmera é enviado ao gerenciador de *closed caption*, que é responsável por encapsular a legenda no vídeo bruto. Após esse processo, o sinal é enviado ao codificador de vídeo responsável por padronizar, melhorar a qualidade e diminuir o tamanho do vídeo.

Assim, o fluxo de vídeo é inserido no Playout, onde o mesmo transcodifica o vídeo para pacotes TS e sincroniza os servidores de SI, EPG, dados, *closed caption*. Toda via o fluxo de pacotes é multiplexado junto ao vídeo H.264 para a transmissão, adotando o padrão brasileiro ISDB-Tb. A seguir na Figura 1 é mostrado o diagrama de implantação.

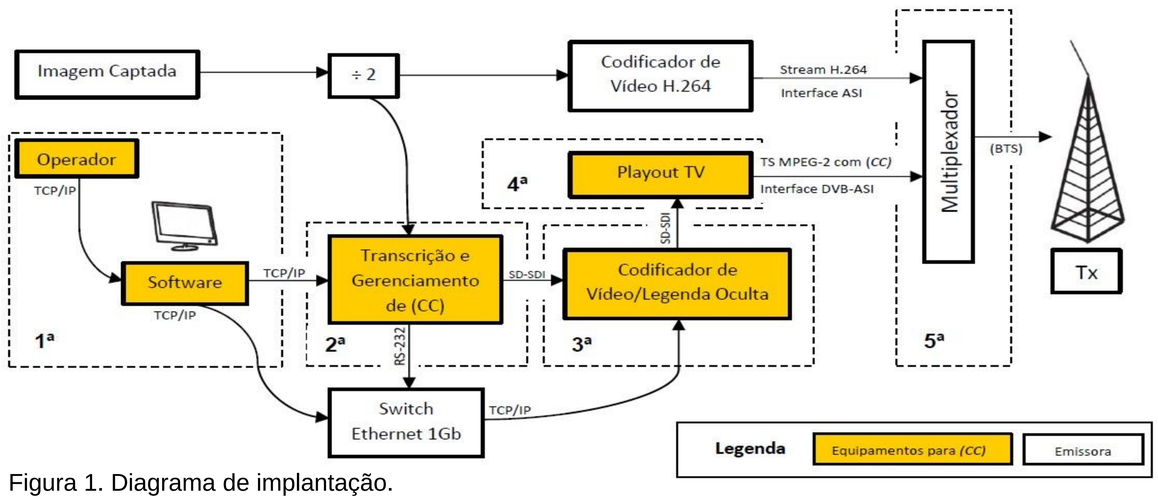


Figura 1. Diagrama de implantação.

O interpretador de legendas exibe os comandos dos caracteres para serem sobrepostos ao vídeo decodificado. Exemplo: “0x1D 0x21” corresponde à nota musical (♪). Os comandos são apresentados na Tabela 1 e Tabela 2.

	0x	1x	2x	3x	4x	5x	6x	7x	8x	9x	Ax	Bx	Cx	Dx	Ex	Fx
x0	NUL		SP	0	@	P	`	p	BKF	COL	100	°	À	Đ	à	õ
x1			!	1	A	Q	a	q	RDF	FLC	¡	±	Á	Ñ	á	ñ
x2			"	2	B	R	b	r	GRF	CDC	¢	²	Â	Ò	â	ò
x3			#	3	C	S	c	s	YLF	POL	£	³	Ã	Ó	ã	ó
x4			$	4	D	T	d	t	BLF	WMM	€	Ž	Ä	Ô	ä	ô
x5			%	5	E	U	e	u	MGF	MACRO	¥	µ	Å	Õ	à	õ
x6		PAPF	&	6	F	V	f	v	ONF		Š	¶	Æ	Ö	æ	ö
x7	BEL		'	7	G	W	g	w	WHF	HLC	§	·	Ç	×	ç	÷
x8	APB	CAN	(	8	H	X	h	x	SSZ	RPC	š	ž	È	Ø	è	ø
x9	APF	SS2	)	9	I	Y	i	y	MSZ	SPL	©	¹	É	Ù	é	ù
Xa	APD		*	:	J	Z	j	z	NSZ	STL	ª	º	Ê	Ú	ê	ú
xB	APU	ESC	+	;	K	[	k	{	SZX	CSI	«	»	Ë	Û	ë	û
xC	CS	APS	,	<	L	\	l	|			¬	Œ	Ì	Ü	ì	ü
xD	APR	SS3	-	=	M	]	m	}		TIME	ÿ	œ	Í	Ý	í	ý
xE	LS1	RS	.	>	N	^	n	~			®	Ÿ	Î	Þ	î	þ
xF	LS0	US	/	?	O	_	o	DEL			¯	¿	Ï	ß	ï	15/15

Tabela 1. Conjunto de comandos dos caracteres latinos.

	0x	1x	2x	3x	4x	5x	6x	7x
x0				¤	…			
x1			♪	¦				
x2				¨	‘			
x3				´	’			
x4				¸	“			
x5				¼	”			
x6				½	•			
x7				¾	™			
x8					⅛			
x9					⅜			
xA					⅝			
xB					⅞			
xC								
xD								
xE								
xF								

Tabela 2. Caracteres especiais.

Resultados e Discussão

Em uma sala isolada do som externo o locutor dita as frases pausadamente para o software de reconhecimento de voz. Na Figura 2 é mostrado o operador ditando as frases.

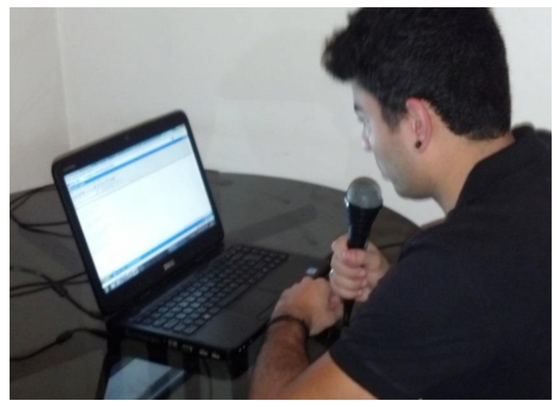


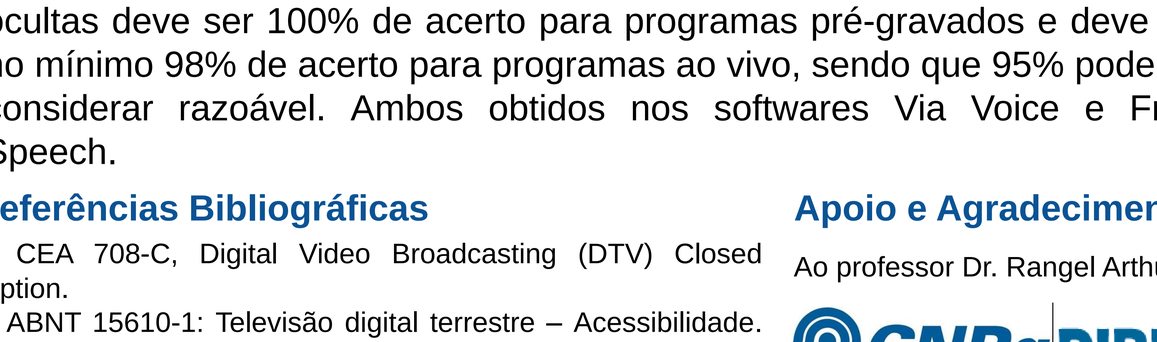
Figura 2. Operador ditando as frases.

Softwares/Quesitos	Via Voice	FreeSpeech	Speech Recognizer
Ano de lançamento	2003	2000	2011
Ano de atualização	2009	2010	2013
Léxico	50.000 palavras	45.000 palavras	Indefinido
Idioma	Português Brasil	Português Brasil	Idiomas do Google Tradutor
Possui correção e adição de palavras ?	SIM	SIM	NÃO
Suporta qual tipo de locutor ?	Dependente	Dependente	Independente
Plataforma	Windows	Windows	Windows/iOS/Linux
Técnica	HMM	HMM	Indefinida

Tabela 3. Comparação entre os Softwares.

A facilidade da interface homem-máquina é uma das características que reduz a carga sobre o operador. A Tabela 3 mostra a comparação entre eles.

Para os testes, foram utilizados dois locutores sendo eles: LD (locutor dependente) com pré-gravação e LI (locutor independente) sem pré-gravação. Foram gravadas e comparadas 60 frases. O gráfico a seguir mostra a porcentagem de acerto dos softwares.



	Free Speech	Via Voice	Speech Recognizer
LI	61,91%	85,35%	87,98%
LD	95,97%	98,30%	0%

Gráfico 1. Taxa de acertos de palavras (%) x listas de frases.

Conclusão

A tecnologia de reconhecimento automático de voz mostrou ser muito eficiente para aplicações de *closed caption*, com ótimo custo-benefício. Essa técnica tem diminuído o custo para as emissoras de TV comparado a técnica de estenotipia. O reconhecimento de voz é mais fácil e barato de ser implementado, pois permite que em 1 semana a empresa já tenha um profissional capacitado. Ao contrario da estenotipia que é mais complexo e varia de seis meses a um ano. Portanto, com os estudos realizados houve resultados satisfatórios alcançando o mínimo de precisão exigido pelo Ministério das Comunicações que estabelece que a produção de legendas ocultas deve ser 100% de acerto para programas pré-gravados e deve ter no mínimo 98% de acerto para programas ao vivo, sendo que 95% pode se considerar razoável. Ambos obtidos nos softwares Via Voice e Free Speech.

Referências Bibliográficas

[1] CEA 708-C, Digital Video Broadcasting (DTV) Closed Caption.

[2] ABNT 15610-1: Televisão digital terrestre – Acessibilidade. Parte 1: Ferramenta de texto.

[3] ARIB STD-B37, Structure and Operation of Closed Caption Data Conveyed Data Conveyed by Ancillary Data Packets.

[4] ARIB STD-B31, Transmission System for Digital Terrestrial Television Broadcasting.

Apoio e Agradecimentos

Ao professor Dr. Rangel Arthur