MODELAMENTO PROSÓDICO PARA CONVERSÃO TEXTO-FALA DO PORTUGUÊS FALADO NO BRASIL

Cairo H. Da Silva

CPqD-TELEBRÁS

Caixa Postal 1579 13088-061 Campinas, SP, Brasil

e-mail:cairo@.cpqd.br. - fone (019) 2396643

Fábio Violaro

UNICAMP-FEE-DECOM

Caixa Postal 6101 13081-970 Campinas, SP,Brasil

e-mail fabio@decom.fee.unicamp.br

Fone (019) 2398324, fax: (019) 2391395

Sumário: Este artigo descreve um modelo de tratamento prosódico aplicado a um sistema de conversão texto-fala para o português falado no Brasil. O modelo prosódico é composto de um modelo duracional e de um modelo de freqüência fundamental. Com base em dados extraídos a partir da análise de realizações de fala natural (frases ditas por pessoas), é proposto um modelo que controla a duração fonética e gera curvas de freqüência fundamental para sentenças declarativas neutras.

Abstract: This paper describes a prosodic model for a text-to-speech synthesis system developed for the portuguese language spoken in Brazil. The model provides both a duration and a fundamental frequency (pitch) control of the phones. The model is based on rules extracted from a corpus of natural speech (phrases spoken by people). These rules control the duration of the phones and define fundamental frequency profiles for neutral declarative sentences.

Palavras chave: Conversão texto-fala, Prosódia, Freqüência Fundamental.

1. INTRODUÇÃO

Um conversor texto-fala é um sistema que aceita como entrada um texto em sua forma ortográfica (isto é, como usualmente se escreve) e produz, como saída, a fala na forma de onda sonora.

Um texto ortográfico é produzido pela combinação das letras que constituem o alfabeto empregado. Já a fala consiste de combinações dos possíveis tipos de sons que o trato vocal humano é capaz de produzir. Estes tipos de sons, que se distinguem suficientemente uns dos outros para serem discernidos pelos ouvintes, são denominados fonemas. Por exemplo, ao ouvir as palavras "pata" e "bata", o ouvinte faz distinção entre elas por meio dos sons de suas respectivas consoantes iniciais. Logo o "p"e o "b" correspondem a sons de diferentes tipos, ou seja, diferentes fonemas. Um fonema é, então, uma categoria de som. A realização física de um fonema é denominada fone ou alofone. Assim os alofones correspondentes à pronúncia de uma mesma palavra por duas pessoas diferentes podem diferir entre si.

O processo de produção de fala a partir de texto consiste basicamente em determinar a seqüência de fonemas referentes ao texto que se queira sintetizar e produzir os fones a eles referentes. Mas isso não é suficiente para produzir uma fala inteligível e natural. Para que a fala sintetizada apresente tais atributos é preciso que no processo de síntese haja um tratamento prosódico.

Mas o que vem a ser prosódia? Para entender este conceito é necessário saber primeiro o que são parâmetros prosódicos. São eles a freqüência fundamental, a duração e a energia. Freqüência fundamental é o nome que se dá à freqüência com

que as cordas vocais de uma pessoa vibram enquanto ela está falando. Esta grandeza é comumente chamada "pitch". Adicionalmente cada fone possui na fala uma duração e um valor de energia.

A prosódia pode ser definida como a evolução dos parâmetros prosódicos ao longo da fala, ou seja, os diferentes valores de pitch, duração e energia que a fala vai assumindo ao longo do tempo.

Dar um tratamento prosódico aos fonemas durante o processo de conversão texto-fala consiste em se determinar a evolução dos parâmetros prosódicos de modo a se obter inteligibilidade e, se possível, naturalidade na fala produzida.

O objetivo do presente trabalho é a construção de um modelo de tratamento prosódico aplicado a um sistema de conversão texto-fala para o português falado no Brasil. Basicamente, o trabalho consiste na análise de realizações de fala natural (frases ditas por pessoas), criação de regras que descrevem o tratamento prosódico a que as pessoas submetem suas próprias falas, uso destas regras para construção de um módulo prosódico aplicado à conversão texto-fala, e emprego do sistema de conversão para testar e aprimorar estas regras.

O modelamento prosódico aqui apresentado leva em conta apenas a duração e o pitch, que são os parâmetros prosódicos mais relevantes.

2. AMBIENTE DE TRABALHO CRIADO

Para construir um modelo prosódico aplicado à conversão texto-fala, é necessário conhecer o "comportamento prosódico" de pelo menos um falante. Por "comportamento prosódico" entenda-se o controle que o falante exerce sobre os parâmetros prosódicos enquanto está falando.

A fim de se conhecer o comportamento prosódico de um determinado falante, inicialmente foram analisadas realizações de 78 sentenças declarativas [2]. Ao pronunciá-las, o falante procurou fazê-lo de forma o mais neutra possível, isto é, sem enfatisar trechos em função de seu significado. Posteriormente, foram analisados enunciados relativos a 86 outras sentenças declarativas pronunciadas por outro falante, também de forma neutra. Na realidade, foram analisados 100 enunciados para cada falante, porém optou-se por descartar alguns em virtude de apresentarem pouca neutralidade. Ambos os falantes são adultos do sexo masculino, sendo que o segundo é um locutor profissional. Os falantes são oriundos dos estados de Minas Gerais e São Paulo, respectivamente. Quanto ao grau de instrução, o falante mineiro possui nível superior completo e o paulista nível secundário. Os enunciados foram digitalizados a uma freqüência de amostragem de 16 KHz com 16 "bits" por amostra.

Para se realizar a análise dos enunciados foi desenvolvido um programa de computador voltado para a análise dos parâmetros prosódicos. Através deste programa pode-se determinar as curvas de freqüência fundamental ao longo de enunciados previamente gravados e as durações de seus fones.

Além de se conhecer o comportamento prosódico de dois falantes, foi preciso também se utilizar de um conversor texto-fala para testar as regras do modelo prosódico durante sua evolução [4]. Utilizou-se um conversor texto-fala concatenativo que emprega a técnica "Pitch Synchronous Overlap and Add" (PSOLA) [3].

3. MARCAÇÃO DE FRONTEIRAS PROSÓDICAS

Qualquer sentença produzida pela língua pode ser submetida a uma análise sintática, onde os constituintes sintáticos são identificados, obtendo-se, assim, a estrutura sintática da sentença. Os constituintes sintáticos são sujeitos, predicados, orações, complementos, etc.

De maneira análoga à análise sintática, pode-se submeter qualquer sentença a uma análise prosódica onde o objetivo é identificar os constituintes prosódicos da sentença que, juntos, compõem a sua estrutura prosódica [4].

Mas o que são constituintes prosódicos? A grosso modo eles podem ser definidos como grupos de palavras adjacentes na sentença, onde cada grupo possui a propriedade de influenciar a evolução dos parâmetros prosódicos ao longo das palavras que o constituem. Por exemplo, a frase "As crianças de rua são o principal problema brasileiro" pode ser dividida em dois constituintes prosódicos: "As crianças de rua" e "são o principal problema brasileiro.". Outro exemplo pode ser a frase "Vieram bastante apressados até o terceiro estágio.", onde "Vieram bastante apressados" e "até o terceiro estágio" correspondem a constituintes prosódicos distintos.

Existe uma estreita relação entre estrutura sintática e estrutura prosódica. De maneira geral, pode-se dizer que cada constituinte prosódico é formado por um ou mais contituintes sintáticos adjacentes, embora seja possível haver parte de um constituinte sintático pertencendo a um constituinte prosódico e o restante pertencendo a outro. Por exemplo, na frase "O problema, infelizmente, da economia brasileira são os maus governantes" tem-se quatro constituintes prosódicos: "O

problema", "infelizmente", "da economia brasileira" e "são os maus governantes". Do ponto de vista sintático, "O problema da economia brasileira" é um só constituinte (sujeito da frase), embora esteja dividido entre dois constituintes prosódicos: "O problema" e "da economia brasileira".

Apesar da relação entre as estruturas sintática e prosódica ser estreita, não existe uma correspondência biunívoca entre elas, ou seja, uma sentença que possui determinada estrutura sintática pode ter várias estruturas prosódicas possíveis.

Mas o quê, então, faz com que um falante escolha produzir um enunciado obedencendo a uma determinada estrutura prosódica e não a outra? A resposta é que esta escolha depende do significado do que ele esteja dizendo (semântica) e do uso específico que ele esteja fazendo das palavras (pragmática).

Dentro do contexto do tratamento prosódico aplicado a um sistema de conversão texto-fala para texto irrestrito, optou-se por derivar a estrutura prosódica da sentença a partir do conhecimento das classes gramaticais das palavras e dos sinais de pontuação. Esta opção deveu-se à dificuldade de se obter informações semânticas e pragmáticas através de processamento automático de texto irrestrito.

No estágio atual, a delimitação dos contituintes prosódicos e, portanto, a determinação da estrutura prosódica, é feita manualmente. Esta limitação deve-se a dois fatores. Primeiramente, no atual estágio do conversor texto-fala em desenvolvimento, ainda não se dispõe de meios automáticos para se determinar as classes gramaticais das palavras constituintes das sentenças. Segundo, uma marcação prosódica manual é, em princípio, isenta de erros. Isto possibilita a concentração de esforços nas etapas posteriores do processamento prosódico. Como exemplo de marcação prosódica, podemos ter "A cotação do ouro ## no mercado paralelo ## sofreu um queda significativa ## nos últimos dezesseis meses.", onde o símbolo "##" é usado para indicar fronteiras entre constituintes prosódicos adjacentes na sentença.

Neste trabalho, a identificação dos constituintes prosódicos é feita por meio da determinação das fronteiras entre eles. Tais fronteiras prosódicas são atribuídas a pontos que coincidem com fronteiras entre constituintes sintáticos.

4. MODELO DE DURAÇÃO

Por modelo duracional aplicado à sintese de fala entende-se qualquer tratamento automático pelo qual as durações dos fones de um enunciado a ser sintetizado possam ser determinadas.

No presente trabalho, o modelo duracional foi construído por meio de regras que determinam a duração de cada fone no momento da síntese. A escolha de se criar um modelo baseado em regras deve-se ao fato desta abordagem não exigir um conjunto de dados de grande dimensão.

Foi criado um modelo cujas regras atuam de forma independente entre si. Esta atuação se dá por meio de um produtório de coeficientes, onde cada regra entra com seu fator. Para cada fone de um enunciado a ser sintetizado é calculado o produtório das regras que incidem sobre ele. O valor obtido para este produtório multiplicado pela "duração média do fone" resulta no valor de duração a ser utilizado em sua síntese.

O modelo pode ser expresso por:

$$D = Dm \times K$$

onde

D é a duração calculada para o fone

Dm é a duração média do fone

K é o valor resultante do produtório de coeficientes.

Mas o que vem a ser essa "duração média do fone"? Para responder a essa pergunta é necessário lembrar que o presente trabalho desenvolveu-se sobre um sintetizador concatenativo. Neste tipo de síntetizador é preciso extrair as unidades concatenativas de realizações de fala natural. Essas unidades para concatenação são constituídas por alofones cujos limites devem ser precisados. Uma vez que se disponha das durações de um conjunto de alofones para cada fone, pode-se, por simples média aritmética, determinar a duração média de cada fone.

Deve-se observar adicionalmente que foram estipulados valores máximos e mínimos para a duração de cada fone. Para estipular estes valores observou-se as durações geradas para um conjunto de frases sintetizadas. As durações de segmentos cuja audição mostrou-se desagradável serviram como referencial para o estabelecimento de limites duracionais. A partir deste referencial, a estipulação de valores exatos para os limites foi feita por tentativa e erro. Se, por exemplo, o valor de duração calculado para um fone ultrapassar a duração máxima estipulada para ele, este valor de duração passará a ser

igual ao valor máximo estipulado. A limitação dos valores de duração para cada fone objetiva eliminar possíveis distorções geradas pelo modelo.

O processo de estabelecimento das regras deste modelo consistiu numa escolha de um subconjunto das regras de um modelo semelhante para a língua inglesa [1]. Foram escolhidas regras que foneticamente pareciam pertinentes também à língua portuguesa. Também foram criadas outras regras específicas para o português. Uma vez estabelecido um conjunto de regras para o modelo duracional, os valores dos coeficientes relativos a elas foram determinados através de sucessivos ajustes orientados pela percepção de sentenças sintetizadas.

As regras atuam segundo a estrutura prosódica de cada sentença e de maneira hierárquica. Assim sendo, existem regras que atuam a nível de sentença, isto é, influenciando todos os fones da sentença. Existem outras regras que atuam a nível de constituinte prosódico de modo a influenciar todos os fones do constituinte. Outras agem a nível de palavra, de sílaba e, por último, de fones. Como exemplo, cita-se a regra abaixo que atua a nível de palavra e tem como efeito provocar uma redução de duração nas palavras longas:

Cada palavra tem a duração média de seus fones multiplicada por um coeficiente segundo o número de sílabas que possui de acordo com a **Tabela 1**. Outro exemplo é a regra, também a nível de palavra, que altera a duração de uma vogal em função do seu contexto direito de acordo com a **Tabela 2**.

Como exemplo de regra que atua a nível de sentença, cita-se a regra abaixo:

Quando uma consoante é precedida por outra consoante, a sua duração é reduzida. O valor do coeficiente desta regra é de 0.83. Também como exemplo de regra que atua a nível de sentença pode-se citar:

São aumentadas as durações dos fones constituintes de palavras que recebem ênfase em função do significado. O valor do coeficiente desta regra é de 1.4. Esta regra ainda não atua no modelo duracional em virtude de ainda não se poder identificar as palavras enfatizadas. Como exemplo de palavra que pode receber ênfase em função do significado, pode-se citar a palavra "corruptos", presente no seguinte enunciado: "Brasília está cheia de corruptos".

Ao todo foram criadas 22 regras. A **Tabela 3** ilustra o resultado da aplicação do modelo duracional sobre a sentença: "O preço da tarifa telefônica foi reduzido.". Os símbolos que constam na coluna "Fone" correspondem à representação fonética da sentença [3].

5. MODELO ENTOACIONAL

Um modelo entoacional aplicado à sintese de fala deve prover tratamento automático pelo qual sejam determinadas curvas de frequência fundamental (F0) para os fones de um enunciado a ser sintetizado. Um modelo de entoação será tão melhor quanto mais se aproximar do comportamento que um falante humano apresentaria. Logo, o caminho natural para se criar um modelo entoacional é baseá-lo no comportameto que a frequência fundamental tem ao longo da fala de uma pessoa. Esta pessoa preferencialmente deve ser a mesma usada para a gravação do dicionário de unidades acústicas caso esteja-se usando um sistema de síntese por concatenação.

Uma vez que sejam extraídos dados sobre a fala da pessoa escolhida, é necessário criar um mecanismo para, através dele, determinar as curvas de freqüência fundamental ao longo dos enunciados que se queira sintetizar. Este mecanismo é o modelo entoacional.

Semelhantemente ao modelo duracional, o modelo entoacional consiste de um conjunto de regras baseadas na análise dos dados coletados, ou seja, essas regras determinam a curva de freqüência fundamental de cada fone no momento da síntese.

Conforme foi dito anteriormente, a construção de um modelo baseado em regras requer um volume de dados relativamente modesto, sendo esta a principal causa da escolha desta abordagem.

Foi criado um modelo entoacional cuja principal característica é basear-se em uma estrutura hierárquia de sentença (Fig. 1). Por esta abordagem, cada nível hierárquico deve obedecer às determinações do nível superior e, por sua vez, gerar determinações para o nível inferior. Para cada enunciado que se queira sintetizar é derivada uma estrutura em forma de árvore. A nível de frase é definido, segundo a modalidade da mesma, um comportamento macroscópico para F0. Note-se, entretanto, que o presente trabalho limitou-se a frases declarativas neutras.

Por comportamento macroscópico para F0 a nível de frase, entenda-se o estabelecimento de estruturas limitantes das variações de F0 a nível dos constituintes prosódicos. Tais estruturas nada mais são do que retas. Para cada constituinte prosódico são criadas duas retas que fazem correspondência entre valores no domínio do tempo e valores de F0. Toda a curva de freqüência fundamental dentro de um constituinte prosódico deve ficar entre os gráficos destas retas. A figura 2 ilustra as retas (AB e CD) geradas para um enunciado composto de um único constituinte prosódico: "_São Paulo".

Dentro de um constituinte prosódico são determinados os valores de F0 entre palavras consecutivas, ou seja, é determinado o valor de freqüência fundamental com que é finalizada uma palavra e iniciada a seguinte. O ponto G da figura 2 ilustra o valor de F0 determinado para a fronteira entre as duas palavras que constituem o enunciado.

Como toda a curva de frequência fundamental dentro de um constituinte prosódico deve ficar entre as retas, fica evidente que a curva de freqüência fundamental de cada palavra deve ficar entre as retas estabelecidas para o constituinte prosódico a que ela pertence.

Dentro de uma palavra são determinados os valores de F0 entre suas sílabas. O ponto H da figura 2 ilustra o valor de F0 determinado para o final da sílaba "PP AH UW" e início da sílaba "LL UW".

De maneira análoga ao que ocorre nos níveis acima, dentro de uma sílaba são determinados os valores de F0 entre os seus fones constituintes.

Neste estágio, cada fone já tem definidos os valores de F0 inicial e terminal, além de ter as retas para orientar a curva de F0 ao longo do fone.

Denomina-se microprosódia o formato que a curva de F0 assume internamente a cada fone. Sendo este um trabalho inicial na área de prosódia, optou-se por deixar para trabalhos posteriores um tratamento mais sofisticado de microprosódia. No presente modelo, os valores de F0 inicial e terminal de cada fone são simplesmente interpolados de maneira linear, isto é, a curva de F0 é uma curva contínua composta por uma sucessão de segmentos de reta, onde cada segmento corresponde a

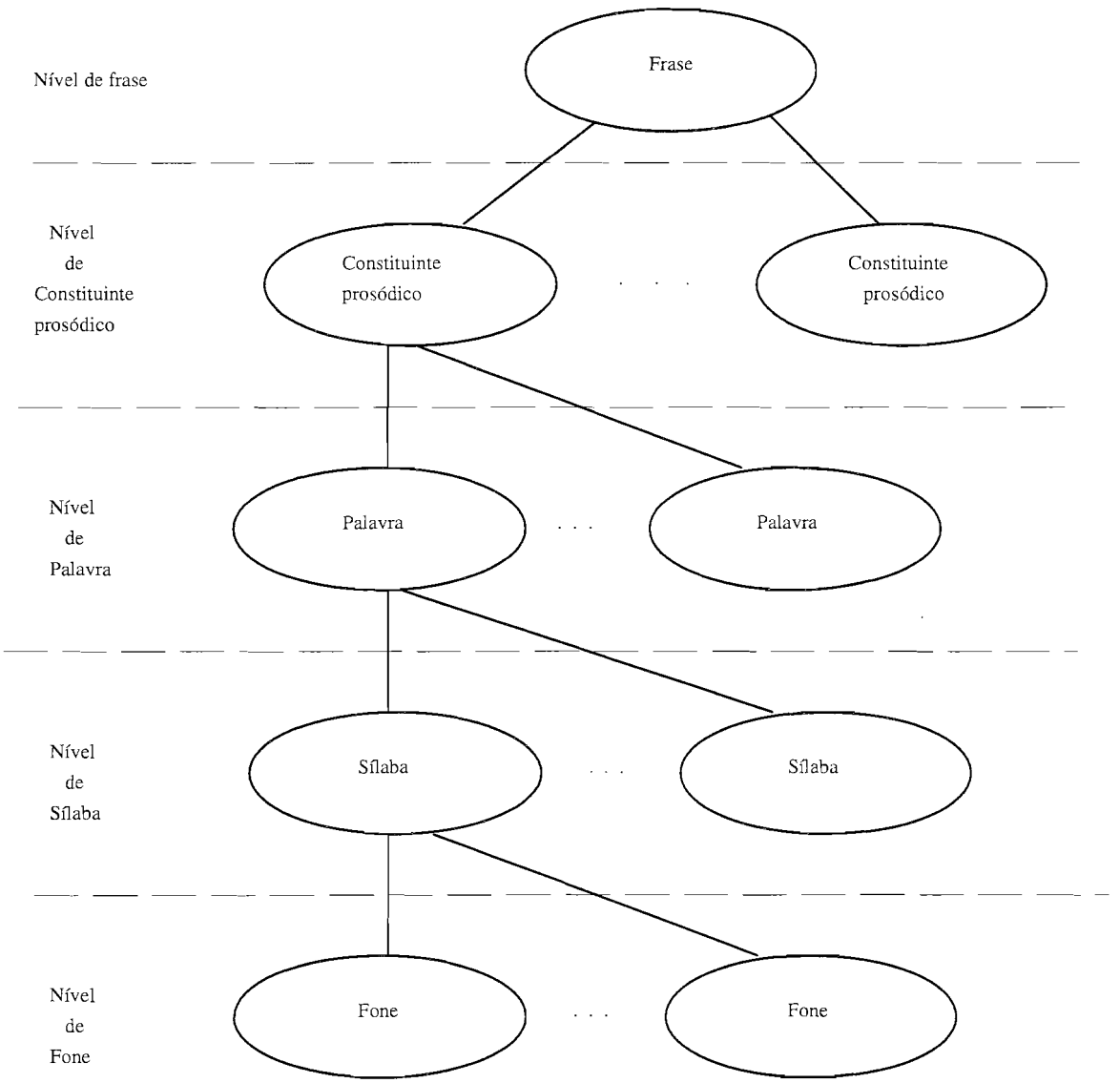

Fig 1 - Árvore de modelamento entoacional.

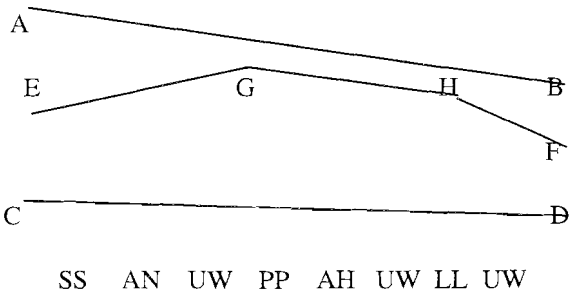

Fig. 2 - Esquema do modelamento entoacional para o enunciado de: "_São Paulo".

uma sílaba. Foi verificado que esta limitação representa, do ponto de vista do ouvinte, uma degradação bastante pequena.

Como exemplo de regra entoacional, cita-se a equação F0 = 100 +2.5* n, onde n é o número de sílabas de um constituinte prosódico correspondente a um sujeito de um período simples. F0 é o valor médio de freqüência fundamental ao final do sujeito, não podendo ser superior a 140Hz. Outro exemplo é a regra que estipula um valor constante para F0 no início de cada sentença.

A **Tabela 4** ilustra o resultado da aplicação do modelo entoacional sobre a sentença: "O preço da tarifa telefônica foi reduzido". As colunas "F0 I" e "F0 D" correspondem, respectivamente, aos valores de freqüência fundamental ao início e ao final dos fones.

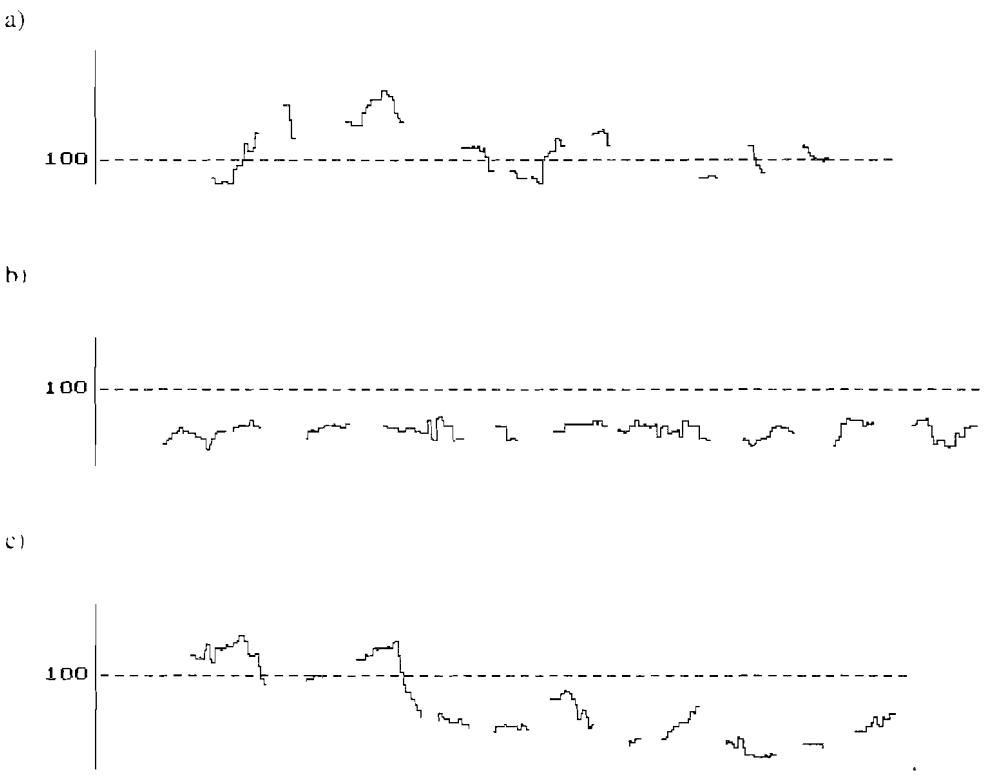

Fig. 3 - Gráficos de F0 para enunciados relativos à sentença "_É necessário que o convênio permita o intercâmbio". (a) F0 natural, (b) F0 intrinseca e (c) F0 modelada

6. AVALIAÇÃO DOS RESULTADOS OBTIDOS

Avaliar os resultados do presente trabalho é uma tarefa difícil. Esta dificuldade deve-se ao fato deste ser um trabalho pioneiro no modelamento da prosódia para a conversão texto-fala, considerando-se que um trabalho desta natureza é dependente do idioma a que ele se aplica. Assim, torna-se impossível fazer uma avaliação por comparação de resultados com um trabalho anterior.

Em termos absolutos, pode-se dizer que o trabalho realizado, apesar de sua simplicidade e caráter inicial, assegurou uma enorme melhoria na qualidade da fala gerada pelo sistema de conversão texto-fala utilizado para teste e avaliação de regras, em comparação à fala sintetizada sem tratamento prosódico.

Para a avaliação do trabalho desenvolvido, recorreu-se ao método de comparação dos parâmetros prosódicos gerados pelos modelos de duração e entonação para uma determinada sentença, com os parâmetros obtidos por meio de uma análise desta mesma sentença quando enunciada por uma pessoa. Este método foi empregado apesar de sua utilidade ser questionável. O fato de uma pessoa poder produzir, para uma mesma sentença, vários enunciados com parâmetros prosódicos bastante distintos torna duvidosa a validade deste método.

A tabela 5 ilustra os valores obtidos para as durações fonéticas pelo modelo de duração desenvolvido (coluna entitulada "Dur Modelada") em comparação com valores obtidos através de análise acústica de um enunciado natural correspondente (coluna entitulada "Dur Natural").

A Fig. 3 mostra gráficos de curvas entoacionais correspondentes a uma realização natural, a uma síntese com freqüência intrínseca (aquela que possuem as unidades do dicionário na forma como foram gravadas) e à saída do modelo entoacional.

7. CONCLUSÕES

Este trabalho descreveu um modelo prosódico que encontra-se em fase de desenvolvimento.

Futuramente, pretende-se ampliar o modelo para tratar, além de sentenças declarativas neutras, sentenças nos modos interrogativo e imperativo. Também pretende-se que a determinação das fronteiras prosódicas das sentenças a serem sintetizadas seja feita automaticamente.

Tabela 1: Controle de duração em função do número de sílabas da palavra

Número de sílabas	Coeficiente	Número de sílabas	Coeficiente
1	1.2	5	0.92
2	1	6	0.9
3	0.97	7	0.9
4	0.94	mais de 7	0.85

Tabela 2: Controle de duração em função do contexto fonético direito da vogal.

contexto fonético direito	coeficiente
fricativa	1.05
oclusiva	1.05
nasal	0.7

Tabela 3: Resultado da aplicação do modelo duracional sobre a sentença: "O preço da tarifa telefônica foi reduzido.".

Fone	Duração (ms)	Fone	Duração (ms)	Fone	Duração (ms)	Fone	Duração (ms)
UW	83	AA	119	FF	126	RR	93
PP	98	RX	40	OO	126	EE	133
RX	40	II	146	NN	85	DD	86
EE	142	FF	126	IY	84	UW	103
SS	131	AA	120	KK	102	ZZ	86
UW	103	TT	85	AA	120	II	146
DD	89	EE	139	FF	145	DD	90
AA	111	LL	81	OO	148	UW	103
TT	85	EE	127	IY	108		

Tabela 4: Resultado da aplicação do modelo entoacional sobre a sentença: "O preço da tarifa telefônica foi reduzido.".

Fone	F0 I (Hz)	F0 D (Hz)	Fone	F0 I (Hz)	F0 D (Hz)	Fone	F0 I (Hz)	F0 D (Hz)	Fone	F0 I (Hz)	F0 D (Hz)
UW	135	131	AA	117	115	FF	109	111	RR	117	112
PP	131	131	RX	115	116	OO	111	112	EE	112	106
RX	131	131	II	116	121	NN	112	107	DD	106	105
EE	131	132	FF	121	115	IY	107	103	UW	105	104
SS	132	125	AA	115	109	KK	103	106	ZZ	104	105
UW	125	121	TT	109	109	AA	106	124	II	105	105
DD	121	119	EE	109	108	FF	124	121	DD	105	100
AA	119	118	LL	108	109	OO	121	119	UW	100	95
TT	118	117	EE	109	109	IY	119	117			

Tabela 5: Comparação entre durações fonêmicas geradas pelo modelo duracional e as correspondentes durações medidas no enunciado do falante paulista utilizado neste trabalho.

Fone	Dur Modelada	Dur Natural
EH	198	78
NN	55	58
EE	64	54
SS	134	90
EE	76	54
SS	149	146
AH	158	134
RX	60	39
IY	40	42
UW	45	20
KK	93	80
IY	64	0
UW	60	68
KK	93	68
ON	124	112
VV	84	60
EE	86	108
NN	50	36
IY	40	0
UW	45	52
PP	102	98
EE	70	70
RX	60	46
MM	74	56
II	137	114
TT	76	68
AA	74	26
UW	60	52
IN	103	68
TT	82	68
EE	72	44
RX	60	46
KK	98	106
AN	153	132
BB	70	62
II	59	72
UW	67	40

AGRADECIMENTOS:

Ao Senhor por ter abençoado este trabalho.

Os autores desejam expressar seus mais sinceros agradecimentos à Prof. Dr. Eleonora C. Albano pela formação básica na área de fonética acústica e a todo o grupo do Laboratório de Fonética Acústica e Psicolingüística Experimental do IEL-UNICAMP.

REFERÊNCIAS:

[1] Allen, J., Hunnicutt, S., & Klatt, D. H.; "From text-to-speech: The MITalk System"; Cambridge, UK, 1987.

[2] Aubergé, V.; "Semi-automatic of a prosodic contour lexicon for text-to-speech system"; Elsevier Science Publishers, no. 39, pp. 274-287, 1992.

[3] Charpentier, F. & Moulines, E.: "Nouvelles techniques de synthèse de la parole"; L'écho des RECHERCHES, no. 137, pp. 37-46, 1989

[4] Egashira, F.; "Síntese de voz a partir de texto para a Língua Portuguesa"; Tese de Mestrado, Faculdade de Engenharia Elétrica da UNICAMP, julho de 1992.

[5] Quené, H. & René, K.; "The derivation of prosody for text-to-speech from prosodic sentence structure"; Computer Speech and Language, no. 6, pp. 77-99, 1992

Cairo Humberto da Silva nasceu em Uberlândia (MG) em 13 de outubro de 1969. Graduou-se em Ciência da Computação pela Universidade Federal de Uberlândia em fevereiro de 1993 e obteve o título de Mestre em Engenharia Elétrica em dezembro de 1995 pela Universidade Estadual de Campinas (UNICAMP). Atualmente, trabalha como pesquisador no Centro de Pesquisas e Desenvolvimento da TELEBRAS (DDS/DTPI/SPSF), desenvolvendo pesquisa na área de conversão texto-fala. Também é aluno de doutoramento da UNICAMP.

Fábio Violaro nasceu em Campinas (SP) em 8 de dezembro de 1950. Possui Graduação, Mestrado e Doutorado em Engenharia Elétrica pela Faculdade de Engenharia Elétrica (FEE) da Universidade Estadual de Campinas (UNICAMP) em 1973, 1975 e 1980 respectivamente. Atualmente é professor do Departamento de Comunicações da FEE/UNICAMP e atua nas áreas de codificação, reconhecimento e síntese de fala.