

Como funcionam alguns fonemas no aplicativo Balabolka

How some phonemes in application Balabolka works

*Edio Roberto Manfio **

**Universidade Estadual de Londrina*

Resumo: O salto gigantesco dado pela tecnologia na área da computação a partir da segunda metade do século vinte nos possibilitou avanços consideráveis nas áreas de *síntese de voz*, *reconhecimento de voz*, *autenticação de voz* ou *comandos e controle por voz*, correspondentes às quatro mais importantes áreas de processamento de sinais da fala. Para os linguistas isso pode significar cada vez mais possibilidades de aplicar mais pontualmente alguns conhecimentos que outrora ficariam restritos apenas aos congressistas, pesquisadores e estudiosos da área. O objetivo desse artigo é justamente comentar, à luz de algumas teorias envolvidas com a Sociolinguística, Geografia Linguística e Dialetoлогия entre outras, sobre ao menos a *síntese de voz*, que é apenas uma das quatro áreas citadas e, no âmbito específico da *síntese de voz*, acerca de um desses vários aplicativos disponíveis: o *Balabolka*.

Palavras-chave: Síntese de voz. Sociolinguística. Geografia Linguística. *Balabolka*.

Abstract: The big leap given by technology in the computing area from the second half of the twentieth century has made possible considerable advances in the fields of *speech synthesis*, *voice recognition*, *voice authentication* or *commands and control by voice*, corresponding to the four most important areas of speech signal processing. To linguists this may mean increasingly more possibilities to apply more promptly some knowledge that once would be restricted only to congressmen, researchers and scholars in the field. The purpose of this article is precisely to comment, in light of some theories involved in Sociolinguistics, Linguistic Geography and Dialetoлогия among others, on at least the voice synthesis, that is only one of the four areas above and within the specific *speech synthesis*, about one of these multiple applications available: the *Balabolka*.

Keywords: Speech synthesis. Sociolinguistics. Linguistic Geography. *Balabolka*.

Introdução

A tecnologia na área da computação certamente deu um salto gigantesco a partir da segunda metade do século vinte. Depois da invenção do transistor e de sua progressiva miniaturização, foi possível criar os microprocessadores que estão cada vez mais rápidos e menores. Em função disso, alguns sonhos da ficção científica tornaram-se realidade ainda no final do século passado: o videofone a baixo custo e em tempo real e o computador que fala, tal como podem ser apreciados no clássico de Stanley Kubrick *2001 Uma Odisseia no Espaço* de 1968, parecem já estar por aí há algum tempo e, ainda que não tenhamos carros voadores como em *Os Jetsons*, os autômatos-empregadas deste mesmo desenho animado já são realidade em países mais desenvolvidos, mesmo que no modo protótipo.

O computador que fala da atualidade, no entanto, merece ressalvas. O Hal 9000, do filme de Kubrick, capaz de interagir verbalmente sobre muitos assuntos com os tripulantes da nave, é tão sofisticado que além dessa tarefa cotidianamente humana, também era responsável pelo funcionamento de todos dos dispositivos da nave, inclusive o controle das vidas de outros tripulantes em hibernação. Ocorre, porém, que ele adquire vontade própria e acaba por colocar toda a missão em xeque. Os nossos computadores pessoais da atualidade ainda não fazem tudo isso.

Em termos de softwares comerciais, sejam eles proprietários ou livres, o que se tem no mercado são alguns poucos e cada qual com recursos bastante limitados se comparados à boa e velha ficção científica. Porém, desconsiderando-se a farta imaginação e criatividade dos livros e do cinema de ficção, esses poucos softwares (algumas dezenas) costumam atender às funções para as quais foram criados – não fosse isso seriam simplesmente descontinuados. Expressões como *síntese de voz*, *reconhecimento de voz*, *autenticação de voz* ou *comandos e controle por voz* são realidade – constituem as quatro áreas de processamento de sinais de voz (HUGO, 1996; DAMASCENO, 2006; TEVAH, 2007; YNOGUTI, 2000) – e tais recursos já funcionam relativamente bem considerando-se as aplicações para as quais se destinam.

A evolução, no entanto, não para e mesmo os softwares que operam apenas com a *síntese de voz* mudam constantemente. Assim, o objetivo desse artigo é justamente comentar, à luz de algumas teorias envolvidas com a Sociolinguística, Geografia Linguística e Dialetoлогия – entre outras sobre o funcionamento de ao menos um desses vários softwares disponíveis: o *Balabolka*.

A metodologia utilizada é bastante simples em função da natureza das análises que são notadamente breves. Foram avaliados alguns sons e prosódias de palavras, frases e fragmentos retirados desse mesmo artigo, todos reproduzidos pelo *Balabolka*. As palavras utilizadas são provenientes de exercícios e exemplos presentes em livros sobre fonética e fonologia – mormente Silva (1999) – e as frases tem como ponto de partida algumas dessas mesmas palavras e exercícios de outras obras como Lopes (1999) e Cagliari & Cagliari (2001, in MUSSALIM & BENTES, 2001).

2 Um pouco sobre o *Balabolka*

Antes de começar a falar sobre o *Balabolka*, é importante delimitar ao menos um termo que será utilizado para fazer referência à natureza desses tipos de softwares. Portanto, para o bem da melhor utilização terminológica – em um movimento contrário ao da simplificação – não chamaremos o *Balabolka* – e outros de natureza similar que porventura sejam mencionados ou avaliados – de software a partir de agora. Utilizaremos o termo *aplicativo*, primeiramente porque software é um termo bastante geral que abarca muitas modalidades e categorias. Depois porque *programa*, ainda que bastante usual no cotidiano, pode conotar diretamente as linguagens de programação que são acessíveis e manipuláveis apenas a iniciados no ramo de programação.

Assim sendo, o *Balabolka* é um aplicativo comercial, ou seja, não é de uso restrito (militar ou institucional). Também pode ser baixado diretamente do sítio oficial que oferece suporte e atualizações (BALABOLKA, 2012). Outra característica é que ele enquadra-se como aplicativo de *síntese de voz*, ou seja, utiliza-se de partes gravadas da voz humana¹ para formar as palavras que pronuncia.

Tal tipo de aplicativo é conhecido como *text-to-speech* ou simplesmente TTS. Aqui o chamaremos de *texto-fala* justamente porque é isso mesmo o que ele faz: transforma o texto escrito em fala. Outros aplicativos com características em comum como o *Dosvox* ou o *IBM Via Voice* também trazem o recurso TTS como o *Balabolka*, porém possuem outras aplicações e especificidades. Estes aplicativos serão discutidos em diferente artigo em outra oportunidade.

Portanto, não fica difícil imaginar para que serve o *Balabolka*. Basta inserir um texto qualquer, escolher uma das vozes disponíveis para idioma que se deseja e ele reproduz as palavras com uma presteza admirável, se comparado, é claro a outros aplicativos com voz nitidamente mais robótica.

É válido lembrar, no entanto, que o aplicativo não faz tradução. Para que ele reproduza em português é necessário que o texto esteja em português e que a voz escolhida – homem ou mulher – não só esteja disponível em sua biblioteca mas também que seja aquela desenvolvida para o português.

Estando, porém, o aplicativo configurado, as atualizações requeridas todas instaladas e o sistema de áudio do computador funcionando adequadamente, ele de fato desempenha bem sua função. Se os dispositivos que funcionam por comando de voz permitem ao usuário usufruir do *hands free*, ou seja, fica com as mãos livres, o *Balabolka* certamente contribui com algo como o *eyes free*, não é mais necessário ler o texto, basta ouvir. É válido lembrar que sistemas *hands free* já existem há um bom tempo: várias montadoras equipam os modelos mais sofisticados de automóveis com

¹ Os diferentes softwares – cada qual baseado em uma linguagem de programação diferente – fazem o uso de maneiras diversas de combinar sons. Os difones e trifones são conceitos bastante comuns entre programadores e consistem em partes da fala humana recortadas com critérios diferentes: ora final de vogal com início de consoante, ora final de consoante com início de vogal, ora partes maiores (TEVAH, 2007, YNOGUTI, 2000).

esse recurso, que pode ser desde um simples comando para ligar ou desligar algo até um sistema completo de acesso à internet com leitura automática de correios eletrônicos entre outras coisas.

Para o ouvinte mais exigente e, eventualmente apreciador de ficção científica, talvez o *Balabolka* possa deixar a desejar. O Hal 9000 do filme há pouco citado – seu nome é inspirado na expressão em inglês *Heuristically programmed ALgorithmic computer* ou Computador Algorítmico Heuristicamente Programado – certamente parece-se muito mais com um ser humano falando. Considerando-se, no entanto, que no filme é de fato um ser humano – evidentemente – que empresta sua voz ao HAL, tanto na versão original em inglês quanto na dublada em português, vale a pena ao menos dar uma olhada – ou melhor, “ouvida” – no *Felipe*, que é uma das vozes masculinas em português do Brasil disponíveis para instalação.

3 *Balabolka* e a prosódia

Um usuário comum, interessado nas vantagens de se ter alguém - ou alguma coisa – lendo o texto para si certamente verá muitas vantagens no *Balabolka*. Ele pode – bastando para isso alguns pequenos, fáceis e rápidos ajustes – falar mais lentamente, em um tom mais alto, com mais ou menos volume e não vai reclamar que está cansado ou que vai deixar o resto para amanhã – como fazem a mãe, o pai ou os avós que leem histórias para as crianças.

Grosso modo, se por um lado tem o potencial de incentivar uma certa vagabundagem na leitura – lembrando que, sabidamente, ler contribui para assimilação da ortografia, do vocabulário e da grafemática de modo geral - por outro pode auxiliar em muito pessoas que sofrem de pouca visão, labirintite, postura ou outro problema de saúde que as impeçam de ficar horas diante de um monitor de computador. Salvos os problemas de saúde ou a pura preguiça de ler, outro fator que favorece o uso talvez seja o de ficar com os olhos livres, como foi dito anteriormente, o que é bastante útil para profissionais que trabalham com aplicativos de animação, desenho, arquitetura entre outros.

Analisando o funcionamento do *Balabolka* com olhos – e, claro, ouvidos – de linguista, porém, a abordagem é diferente – e os comentários também.

A prosódia é incrivelmente próxima da situação real da fala – como já foi dito, se comparado a outros aplicativos de mesma natureza, é claro. As frases afirmativas, os começos e finais de frase, as frases apositivas entre outras são muito bem gerenciadas pelo aplicativo.

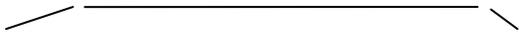
Importante lembrar que, mesmo sendo o português do Brasil uma língua não tonal – como o Mandarim, por exemplo –, os aspectos prosódicos são importantíssimos na distinção entre um modo e outro de pronunciar uma mesma frase. Nas palavras de Lopes (1991, p. 121), “a entonação é um meio importante para diferenciar sentidos a partir da variabilidade da altura da voz durante as emissões

sonoras, e a partir das variações musicais que se apreendem de modo relativo – umas em relação às outras.”

Uma frase como ‘Nós estamos utilizando o *Balabolka* hoje’, por exemplo, é reproduzida de modo a atender os procedimentos prosódicos básicos. Em termos de tonalidade (altura)² há uma certa ênfase ao início e um final mais baixo. As nuances de volume também são controladas pois, não só as sílabas tônicas são obedecidas como também algumas ênfases em certas palavras podem ser percebidas: os pronomes, os artigos, preposições e conjunções são quase sempre pronunciados de modo mais sutil.

Representando graficamente de modo simplificado teríamos no exemplo (a)³:

‘Nós estamos utilizando o *Balabolka* hoje’ (a)



Em se tratando de palavras, expressões ou frases apositivas o aplicativo novamente se sai bem. Com uma frase relativamente curta, dotada de uma oração com valor apositivo, pode-se notar as pausas bem delimitadas no balisamento feito pelas vírgulas:

‘O *Balabolka*, que é um programa de fácil utilização, mostra-se bastante útil a diversas pessoas.’ (b)

Em se tratando de enumeração, ou seja, elementos separados por vírgula seguidos pela conjunção ‘e’ antes do último item, as pausas, volume e tonalidade também são satisfatórios:

‘O *Balabolka* reproduz as pausas, o tom, o volume e as sílabas tônicas.’ (c)

Um fragmento desse mesmo artigo, também dotado de frases com características apositivas foi reproduzido de modo bastante satisfatório, tanto usando

² No cotidiano, em se tratando de som, as pessoas confundem **altura** com **volume**. Independentemente das nomenclaturas utilizadas pelos diferentes teóricos, é válido lembrar que **altura** está relacionada ao tom, ou seja, a frequência das vibrações das cordas vocais perceptível quase que somente quando da pronúncia de vogais. O **volume** diz respeito à pressão sonora: fala normal, sussurrada, aos gritos, vociferada entre outras. Borba (1991) prefere o termo **intensidade** a **volume** e Lopes (1999) utiliza-se do conceito de **tonalidade** a **altura**.

³ Para cada um dos exemplos (a), (b), (c), Quadro 1 e Quadro 2 há um arquivo de áudio salvo pelos recursos do próprio *Balabolka* disponível para consulta. Os endereços eletrônicos estão no **Anexo** deste artigo. Para ouvi-los a partir do **Dropbox**, basta clicar no canto superior direito em ‘download’, depois ‘direct download’ e repetir a operação se for necessário.

vírgulas quanto hífens. Note-se que, neste fragmento, disposto no Quadro 2 há também itens enumerados como em (c):

Um usuário comum, interessado nas vantagens de se ter alguém - ou alguma coisa - lendo o texto para si certamente verá muitas vantagens no <i>Balabolka</i> . Ele pode – bastando para isso alguns pequenos, fáceis e rápidos ajustes – falar mais lentamente, em um tom mais alto, com mais ou menos volume e não vai reclamar que está cansado ou que vai deixar o resto pra amanhã - como fazem a mãe, o pai ou os avós que leem histórias para as crianças.	hífens por vírgulas
Um usuário comum, interessado nas vantagens de se ter alguém, ou alguma coisa, lendo o texto para si certamente verá muitas vantagens no <i>Balabolka</i> . Ele pode, bastando para isso alguns pequenos, fáceis e rápidos ajustes, falar mais lentamente, em um tom mais alto, com mais ou menos volume e não vai reclamar que está cansado ou que vai deixar o resto pra amanhã, como fazem a mãe, o pai ou os avós que leem histórias para as crianças.	vírgulas por hífens

Quadro 1

As perguntas, no entanto, deixam bastante a desejar. Por exigirem muito mais nuances da prosódia, estas não são reproduzidas da forma como utilizamos no cotidiano ou como imaginamos na fala de um personagem literário, por exemplo. Isso pode ser melhor visualizado analisando-se o Quadro 2 a seguir e esse quadro baseia-se em exemplos encontráveis em Lopes (1999) e Cagliari & Cagliari (2001, in MUSSALIM; BENTES, 2001).

<u>Ele virá aqui hoje.</u>	quem	AFIRMAÇÃO
<u>Ele virá <u>aqui</u> hoje.</u>	onde	
<u>Ele virá aqui <u>hoje</u>.</u>	quando	
<u>Ele virá aqui hoje?</u>	quem	PERGUNTA
<u>Ele virá <u>aqui</u> hoje?</u>	onde	
<u>Ele virá aqui <u>hoje</u>?</u>	quando	

Quadro 2

A ênfase individual em cada uma das palavras sublinhadas pode mudar completamente a afirmação ou a pergunta e isso o *Balabolka* não consegue fazer.

É importante ressaltar, no entanto, que a complexidade envolvida no processo de falar cada uma das frases do Quadro 2 é bastante grande e nos faz pensar

no quanto tempo cada um de nós demora para ter destreza linguística suficiente a ponto de dominá-la. Veja-se, por exemplo que crianças até determinadas idades podem não distinguir as nuances como estas apresentadas.

Evidentemente, para reproduzir a maior parte das possibilidades prosódicas utilizadas no cotidiano por falantes do português do Brasil ele teria de possuir um banco de dados muito maior e o computador, onde ele estivesse instalado, grande poder de processamento.

Isso se dá, resumidamente porque, como foi dito anteriormente, o *Balabolka* enquadra-se como aplicativo de *síntese de voz* e, para tanto, utiliza-se de partes gravadas da voz humana para formar as palavras que pronuncia: quanto mais nuances prosódicas, mais partes devem ser gravadas e gerenciadas pelo aplicativo.

4 Sobre as realizações de fala e o registro

O processo de montagem de palavras do *Balabolka* é artificial, ainda que procure reproduzir com a maior presteza possível cada um dos fonemas da língua portuguesa. Após ouvir vários textos a partir dele, começamos a perceber os diferentes graus de artificialidade na cadeia da fala. Fosse uma criança, bastaria dar alguns exemplos e fazê-la repetir algumas vezes. Quanto ao *Balabolka*, a única maneira de corrigi-lo é aumentando seu banco de dados – uma tarefa que também envolve programadores.

As realizações de fala do *Balabolka* – se é que, para o caso de um aplicativo podemos falar em realizações de fala – podem nos remeter a um registro linguístico específico. O *Felipe* – uma das vozes masculinas em português do Brasil disponíveis – não parece ser do nordeste ou extremo sul do Brasil. Também não faz lembrar alguém de nativo de capitais como Cuiabá, Belo Horizonte ou Curitiba. Lembra sim, em uma primeira audição, um apresentador de jornal televisivo de abrangência nacional.

O cuidado em não denominar categoricamente os sons produzidos pelo aplicativo como *realização de fala* pode não se justificar se levarmos em consideração que a montagem feita por ele segue um algoritmo construído por falantes nativos – os programadores – e, portanto, tende a agradar aos ouvidos dos outros ouvintes que igualmente identificam-se com a mesma língua. Assim sendo, os sons, em última análise, são perceptivelmente artificiais.

Os sons do aplicativo, no entanto, como dito há pouco, podem mesmo nos remeter a um registro linguístico específico. Muito do que se encontra em termos de realizações da fala em programas televisivos de grande abrangência também dão essa impressão. Basta, no cotidiano, alguém ter a voz um pouco mais empostada, pronúncia mais apurada e uma certa correção ao colocar pronomes e conjugar verbos para virar chacota de outros colegas por parecer-se com apresentador de telejornal, ator de teatro (novela, cinema) ou narrador futebolístico. Isso ocorre nesses meios porque – entre outras coisas – a fala é sempre mais tensa (BORBA, 1991).

No cotidiano, porém, em que ocorre a fala menos tensa ou frouxa (BORBA, 1991) isso pode soar um tanto artificial (CAMARA JR, 2006; 2004), lembrando mesmo as pronúncias ensaiadas (TV, teatro, cinema), estilísticas (narrações) e didáticas (docentes de ensino fundamental), e é o que ocorre com o *Felipe*.

Resta responder porque então o *Felipe* e os apresentadores de Telejornal (entre outros, como colocado) não só soam artificiais comparados ao uso linguístico cotidiano como também notoriamente não parecem ter ‘origem’ sulista ou nordestina.

Historicamente sabemos que todas as regiões de maior prestígio sócio-político-cultural exercem grande influência sobre as outras menos privilegiadas, em qualquer continente ou país. Em uma importante capital de determinado país, não só a língua ali falada como o registro linguístico em que se manifesta passa a ter certo privilégio em relação à de outras cidades mais pobres ou menos influentes politicamente. Para o caso do português do Brasil, o padrão falado, além de atender a muitos dos preceitos ditados pelas gramáticas, fundamenta-se a partir do padrão utilizado pelas classes cultas que se concentram nessas capitais (BORBA, 1991; CAMACHO, 2001; LOPES, 1999; SILVA, 1999).

O registro encontrado no eixo Rio-São Paulo parece há muito tempo influenciar o modo como as maiores mídias operam com uma espécie de padronização dialetal, provavelmente por há muito serem tidas como grandes centros irradiadores de cultura (BORBA, 1991). Basta verificar o quanto é raro – ou quase impossível, para sermos mais francos e empíricos – ouvirmos um repórter, ator ou apresentador dos grandes canais de TV utilizando um ‘erre’ retroflexo quando do exercício de sua função. Não fica, portanto, difícil de verificar o motivo pelo qual o *Felipe* pode fazer lembrar o Bonner, o Heródoto ou o Cid⁴.

O processo de melhoramento de um aplicativo como este, portanto, não se restringe apenas a fazê-lo funcionar nas diferentes versões dos sistemas operacionais mais conhecidos (Windows, Linux ou OS). Além das diferentes plataformas (sistemas operacionais) devem também operar nas configurações dos computadores mais modernos e ainda atender aos quesitos de usabilidade e ambiência amigável. Atendidas todas essas necessidades, resta torná-lo mais convincente aos ouvidos.

O *Felipe*, que não é a única voz disponível – há a *Juliana* e o *Alexandre*, muito inferiores em qualidade – já é uma evolução: a voz é mais empostada, a montagem é mais próxima à da cadeia da fala, a qualidade do som é superior. Para melhorar, lembrando que outros aplicativos proprietários já são superiores, a nova voz teria de ser menos susceptível a críticas, ou seja, as análises prosódicas e fonético/fonológicas precisam apontar menos discrepâncias em relação a um locutor humano, independente da região à qual ele pertence.

Para tanto, apresentamos a seguir mais uma breve análise, desta vez, voltada a alguns sons do português do Brasil.

⁴ William Bonner, Heródoto Barbeiro e Cid Moreira, entre outros, estão entre os mais conhecidos apresentadores de telejornais do Brasil.

5 Enfim, alguns dos fonemas

Há muitos fonemas que poderiam ser analisados no *Balabolka*, porém, o espaço destinado para esse artigo não permitiria. Começemos, portanto, com os sons do ‘erre’, que se realiza de diversas maneiras em uma mesma palavra e de modo diverso nas regiões do Brasil.

Um dos motivos para a escolha do ‘erre’ é que, no Brasil, ainda que tenha fonologicamente falando poucos valores distintivos se comparados às várias possibilidades de realização em termos fonéticos, constitui muitas vezes motivo para preconceito linguístico, principalmente quando se trata do contraste capital/interior.

Para as transcrições aqui realizadas, foi feito o uso da fonte SILDoulos IPA93, disponível gratuitamente na página do *Summer Institute of Linguistics - SIL International* (2012) e em vários sites relacionados às disciplinas de Fonética, Fonologia, Sociolinguística, Geografia Linguística, Dialectologia entre outras.

As três posições do ‘erre’ mais analisadas e encontráveis em palavras do português brasileiro são: começo, meio e fim (SILVA, 1999; BORBA, 1991; CAGLIARI & CAGLIARI, 2001; AGUILERA, 1990; CALLOU e LEITE, 2001).

Em meio de palavra, o *Felipe* parece utilizar-se do tepe, como em quase todo o Brasil. Palavras como arara, marajá, prata, graxa, brava, cara, barata, parada (SILVA, 1999, p. 48) e outras como barato, carango, caro, caroteno, maracujá já servem para verificar qual é o fonema e seu ponto de articulação⁵. Nem sempre o tepe tem boa qualidade e isso é verificável quando mantém algumas vizinhanças. A análise sistemática do problema não será feita aqui primeiramente por questão de espaço, depois por não se tratar exatamente de ponto de articulação bucal, mas do modo como os sons foram concatenados.

Em começo de palavra, o som é nitidamente velar [x] ou glotal [h] – ainda que, evidentemente, esteja aberto a discussões -, independente da vogal que o segue. As palavras rata, rapaz, rama, rala (SILVA, 1999, p. 49) e outras como rato, remo, rede, roda, rima, runa foram utilizadas como exemplo.

Em final de palavra, algumas considerações extras precisam ser feitas. Ainda que o ponto de articulação seja claramente alveolar, como no tepe, há uma pequena fricção, muito utilizado pelos apresentadores de Telejornal. O tepe realiza-se e é seguido por uma fricção criada pela passagem do ar entre a língua e a região alveolar. Está mais próxima de uma fricativa lateral, tal como pode ser verificado na tabela do alfabeto internacional de fonética (SILVA, 1999, p. 41) ou em comentário feito por Borba (1991, p. 121). Essa realização, no entanto, só ocorre quando a palavra é pronunciada separadamente. Quando em contexto, momento em que o ‘erre’ mantém vizinhança com outros fonemas, a realização é diferente.

⁵ Para todos esses exemplos também há um arquivo de áudio salvo pelos recursos do próprio *Balabolka* disponível para consulta. Os endereços eletrônicos estão no Anexo deste artigo.

A palavra ‘mar’, por exemplo, quando pronunciada em separado realiza-se [maɾ], tal como descrito anteriormente como uma fricativa lateral. Essa mesma palavra, quando pronunciada em contexto como na frase ‘o mar está calmo’, manifesta-se com um tepe propriamente dito, ou seja, [mar]. Esse detalhe não nos dá ainda margem para afirmar que o locutor humano utilizado na coleta do banco de sons para o *Felipe* seja paulistano, por exemplo. Primeiro porque, ainda que a maioria dos paulistanos natos façam notoriamente o uso do tepe em final de palavras como [mar], não realizam na fala cotidiana [maɾ], com fricção. Depois porque, mesmo no oeste paulista, onde ‘mar’ realiza-se [maɾ], sua realização em contexto como ‘o mar está calmo’ torna-se também [mar] ou, em transcrição completa, [umaristakawmɔ].

De acordo com Silva (1999), os sons tendem a ser modificados pelo ambiente em que se encontram. Trata-se, portanto, da primeira premissa que deve ser levada em consideração quando se trata de transcrição fonética. Palavras como *desdém* ou *vesga* são também bons exemplos para isso: [dezdeɪ] e não [desdeɪ].

A propósito, o *Felipe* faz uma boa distinção também destes tipos de visinhança fônica, ou seja, no interior de uma mesma palavra. Nas frases ‘Ele não comia **desde** ontem’ e ‘Ele não comia **deste** ontem’, as palavras *desde* e *deste* realizam-se de modos diferentes bem ao gosto dos ouvidos de falantes nativos de língua portuguesa do Brasil. O aplicativo gerencia adequadamente o fato de, no primeiro caso, a fricativa [s] avizinhar-se da oclusiva **vozeada** [d] e ter de ser realizada também com vozeamento em [z]. Da mesma maneira faz a adequação quando a mesma fricativa [s] avizinha-se da oclusiva **desvozeada** [t], mantendo a igualmente desvozeada [s].

Note-se por essas diversificações de pronúncia dentro ou fora do contexto que o *Balabolka* de fato surpreende quanto à funcionalidade. Em programas como *Dosvox* ou *IBM Via Voice* isso funciona de modo um pouco diverso, como poderá ser verificado em outra oportunidade.

Muitos outros sons do *Felipe* no *Balabolka* podem ser comentados e discutidos. Quanto aos ditongos, por exemplo, o *Balabolka* deixa bastante a desejar. As vogais frouxas como [ɪ] e [ʊ], muito recorrentes em algumas palavras como *árdua*, *mágoa*, *tênue*, *côngrue* (SILVA, 1999) são pronunciadas separadamente e soam muito artificiais - a qualquer ouvido.

6 Considerações finais

A *síntese de voz*, ainda que tenha evoluído muito, não passa despercebida aos linguistas de modo geral. Há vários relatos curiosos do dia-a-dia em que, por exemplo, algumas pessoas chegam ao ponto de dizer obrigado ao ‘robô’ que fala do

outro lado da linha. Isso ocorre porque a URA, ou Unidade de Resposta Audível⁶, um dos recursos muito comuns na telefonia mundial, opera com um sistema de *síntese de voz* muito mais apurado que o *Balabolka* entre outros. A explicação é simples: quanto mais próximo da pronúncia humana estiver a voz do computador, mais acentuam-se os conceitos de usabilidade e ambiente amigável. Em outras palavras, a irritação do cliente pode ser minimizada uma vez que ele consegue ao menos entender o que está sendo dito.

Nesse sentido, vale a pena avaliar o funcionamento do *Balabolka* - e outros como o *IBM via Voice* ou *Dosvox* em outras oportunidades - justamente para se ter um parâmetro mínimo de comparação, tendo em vista que comparar diretamente seu desempenho com o de um ser humano em termos de realização de fala é ainda um tanto desfuncional. O Hall 9000 de Kubrick ainda é ficção, mas as pesquisas estão evoluindo e nossa contribuição, ainda que pequena, certamente auxiliará na construção de dispositivos mais sofisticados e, de fato, úteis para o ser humano.

Referências

2001, UMA ODISSEIA no Espaço. Direção: Stanley Kubrick. Produção: Stanley Kubrick. Intérpretes: Keir Dullea, Gary Lockwood, William Sylvester, Douglas Rain e outros. Roteiro: Stanley Kubrick e **Arthur C. Clarke**. Música: **Johann Strauss II** e **Richard Strauss**. 1968. Califórnia.

AGUILERA, V. A. *Atlas Linguístico do Paraná*. 1990. Tese (Doutorado em Letras) – Universidade Estadual Paulista Júlio de Mesquita Filho, Assis. 1990. 2v.

ALKMIM, Tânia. *Sociolinguística* (parte I) in MUSSALIM, Fernanda; BENTES, Anna Christina (Org.). *Introdução à linguística I: domínios e fronteiras*, v.1. São Paulo: Cortez, 2001.

BALABOLKA. Disponível em: <<http://www.cross-plus-a.com/br/balabolka.htm>>. Acesso em: 13 jan. 2012.

BORBA, Francisco da Silva. *Introdução aos estudos linguísticos*. 11 ed. Campinas, SP: Pontes, 1991.

CALLOU, D.; LEITE, Y. *Iniciação à fonética e à fonologia*. Rio de Janeiro: Jorge Zahar, 2001.

CAMACHO, Roberto Gomes. *Sociolinguística* (parte II) in MUSSALIM, Fernanda; BENTES, Anna Christina (Org.). *Introdução à linguística I: domínios e fronteiras*, v.1. São Paulo: Cortez, 2001.

⁶ Um dos sistemas que opera com a URA no Brasil é o Asterisk que, em tempo oportuno, merece também atenção dos linguistas.

CÂMARA JR, Joaquim Mattoso. *Estrutura da língua portuguesa*. 13. ed. Petrópolis, RJ: Vozes, 2004.

CÂMARA Jr, Joaquim Mattoso. *Para o estudo da fonêmica Portuguesa*. Rio de Janeiro: Organização Simões, 2006.

DAMASCENO, Eduardo Filgueiras. *Avaliação das bibliotecas de reconhecimento e síntese de fala em ambientes virtuais*. 2005. 129 f. Dissertação (Mestrado em Ciência da Computação) – Centro Universitário Eurípedes de Marília - UNIVEM. Marília, 2006. Disponível em: <http://www.fundanet.br/servico/aplicativos/mestrado_dir/dissertacoes/Avalia%E7%E3o_da_s_bibliotecas_de_reconhecimento_e_s%E9ntese_de_f_52_pt.pdf>. Acesso em 12 fev. 2012.

DAMASCENO, Eduardo Filgueiras; PEREIRA, Tatiane Valau, BREGA, José Remo Ferreira. *Implementação de serviços de voz em ambientes virtuais*. Artigo. Disponível em: <<http://www.dcc.ufla.br/infocomp/artigos/v4.3/art09.pdf>>. Acesso em 12 fev. 2012.

HUGO, Marcel. *Uma interface de reconhecimento de voz para o sistema de gerenciamento de central de informação de fretes*. 1995. 60 f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Estadual de Santa Catarina. 1996. Disponível em: <<http://www.tede.ufsc.br/teses/PEPS0413-D.pdf>>. Acesso em 14 fev. 2012.

LOPES, Edward. *Fundamentos da linguística contemporânea*. 17 ed. São Paulo: Cultrix, 1999.

MUSSALIM, Fernanda; BENTES, Anna Christina (Org.). *Introdução à linguística I: domínios e fronteiras*, v.1. São Paulo: Cortez, 2001a.

MUSSALIM, Fernanda, BENTES, Anna Christina (Org.). *Introdução à linguística II: domínios e fronteiras*, v.1. São Paulo: Cortez, 2001b.

OS JETSONS (The Jetsons). Série de desenho animado. Direção: William Hanna e Joseph Barbera. Produção: William Hanna e Joseph Barbera. Estados Unidos. 1962.

SIL INTERNATIONAL. Disponível em: <<http://www.sil.org/>>. Acesso em: 03 jun. 2012.

SILVA, Thaís Cristófar. *Fonética e Fonologia do Português: roteiro de estudos e guia de exercícios*. São Paulo: Contexto, 1999.

TEVAH, Rafael Teruszkin. *Implementação de um sistema de reconhecimento de fala contínua com amplo vocabulário para o português brasileiro*. 2006. 91 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2007. Disponível em: <<http://www.pee.ufrj.br/teses/textocompleto/2006053001.pdf>>. Acesso em 22 mar. 2012.

YNOGUTI, Carlos Alberto. *Reconhecimento de Fala Contínua Usando Modelos Ocultos de Markov*. 1999. 138 f. Tese (Doutorado em Engenharia Elétrica) – Unicamp. Campinas, 2000. Disponível em: <http://www.decom.fee.unicamp.br/lpdf/teses_pdf/Tese-Doutorado-Carlos_Alberto_Ynoguti.pdf>. Acesso em 16 mar. 2012.

Anexo

Endereços eletrônicos dos arquivos de áudio

I- ‘Nós estamos utilizando o Balabolka hoje’:

<https://www.dropbox.com/s/0kn3zpixhmljcmv/n%C3%B3s%20estamos.wav?m>

II- ‘O *Balabolka*, que é um programa de fácil utilização, mostra-se bastante útil a diversas pessoas’:

<https://www.dropbox.com/s/0p0vrcbtai65fts/o%20Balabolka.wav?m>

III- ‘Um usuário comum (...)’:

<https://www.dropbox.com/s/5fnzticqivxte1t/um%20usu%C3%A1rio%20comum.wav?m>

IV- ‘Ele virá aqui hoje’:

<https://www.dropbox.com/s/uuoagk7vxjci5kx/Ele%20vir%C3%A1%20aqui%20hoje%20%282x%29.wav?m>

V- Arara (...):

<https://www.dropbox.com/s/bv0wcrub0ugr1i7/arara.wav?m>

VI- Barato (...):

<https://www.dropbox.com/s/urnb6rigtyv1o3z/barato.wav?m>

VII- Rata (...):

<https://www.dropbox.com/s/xdrx8zp3yyhevu3/rata.wav?m>

VIII- Celular (...):

<https://www.dropbox.com/s/66hwy9s5bgs8sf4/celular.wav?m>

VIII- ‘O mar está calmo’:

[https://www.dropbox.com/s/yavjpdumzorzqim/o%20mar%20est%C3%A1%20calmo.wav?](https://www.dropbox.com/s/yavjpdumzorzqim/o%20mar%20est%C3%A1%20calmo.wav?m)
m

Recebido em 15 de novembro de 2012.

EDIO ROBERTO MANFIO

Aluno especial em Estudos da Linguagem Universidade Estadual de Londrina (UEL). Mestre pela Universidade Estadual de Maringá (UEM). Docente das Fatecs de Presidente Prudente e Ourinhos. E-mail: edio@femanet.com.br.