TEMA

Do texto à fala

LUÍS CALDAS DE OLIVEIRA*

Esta palestra pretende descrever os processos necessários para converter texto escrito em fala e o desenvolvimento desta tecnologia para a Língua Portuguesa realizado num trabalho de colaboração entre o INESC ID e o CLUL (Centro de Linguística da Universidade de Lisboa). O objectivo é a criação de um sistema que se denomina "sintetizador de fala a partir de texto" e que não é mais do que uma máquina de leitura. Uma vez que, uma grande parte dos sistemas informáticos apresentam os seus resultados na forma de texto, a utilização desta máquina pode dar voz ao computador. Esta facilidade conjugada com a capacidade de reconhecer o que é dito permite a realização de sistemas em que o utilizador dialoga com a máquina.

Apesar da sequência de procedimentos para a conversão de texto em fala ser mais ou menos a mesma nas diferentes Línguas, cada um desses passos varia grandemente em função da Língua em questão. Deste modo, a disponibilidade de um sistema numa dada Língua não significa que possa ser facilmente adaptado para funcionar adequadamente noutra Língua.

Quais são, então, as etapas nesta viagem do texto até à fala? Em primeiro lugar, necessitamos de normalizar o texto de entrada para substituir os símbolos que misturamos habitualmente com o texto: números, datas, horas, quantias em dinheiro, barras, operadores aritméticos, etc.

O passo seguinte é a conversão do texto já normalizado para uma representação mais próxima da oralidade, uma vez que a forma ortográfica é, muitas vezes, ambígua. Por exemplo, na palavra "colo" a letra "o" corresponde a dois sons diferentes da mesma letra. Na transcrição fonética do texto, cada símbolo corresponde apenas a um som (por exemplo kOlu). Outros problemas mais complicados surgem com as palavra homógrafas ("o almoço", "eu almoço", "cheio de sede", "a sede do clube", etc.).

R-100, ROBOT CONSTRUÍDO PELO CENTRO DE KAWASAKI

Os sons caracterizam-se também pela sua duração e timbre. É preciso atribuir uma entoação ao enunciado que depende em grande medida da função das palavras na frase. Este é um processo complicado porque a máquina não sabe o que diz, mas o ser humano que a está a ouvir deverá compreender o sentido da frase. A máquina precisa de informação sobre o valor sintáctico de cada palavra e, com a ajuda da pontuação, atribui a entoação e a duração adequada.

O passo final consiste em criar um sinal de fala com a sequência de sons, duração e timbre anteriormente determinados. Actualmente, o processo mais comum para a realização desta última etapa consiste na utilização de um inventário de segmentos de fala gravados por um orador humano. A máquina justapõe os segmentos necessários para a criação da frase pretendida e, se necessário, modifica as características de duração e timbre para os valores adequados. Estes segmentos têm, em geral, uma dimensão muito reduzida, consistindo normalmente na transição entre dois sons. Considerando que temos aproximadamente 40 sons distintos em Português Europeu, necessitamos de 40x40=1600 segmentos para cobrir todas as transições entre dois sons. O resultado da concatenação destes segmentos é uma voz que é semelhante à do orador original mas dizendo algo que ele, provavelmente, nunca disse.

Geralmente, o objectivo é a leitura de qualquer texto dessa Língua. Dada a variabilidade da Língua, surgirão sempre junções de concatenação com problemas ou erros na atribuição da entoação ou na estimativa da duração dos sons. Estes problemas podem ser minimizados e a qualidade da fala sintética melhorada se restringirmos o vocabulário disponível. Um exemplo de uma solução para um vocabulário limitado é o sistema que foi desenvolvido para a leitura de números de telefones do sistema informativo 118 da Portugal Telecom. Neste caso, a entoação foi definida pelo agrupamento dos dígitos do número e as coarticulações entre sons são conhecidas "a priori" permitindo uma qualidade muito próxima da natural.

Outra área que se destaca no domínio das aplicações da síntese de fala é a dos auxiliares de comunicação a pessoas com necessidades especiais. O sintetizador pode ajudar pessoas com limitações visuais na utilização de computadores pessoais ou de máquinas, como as caixas multibanco ou os dispensadores de bilhetes.

Uma aplicação desenvolvida em colaboração entre o CLUL e o INESC, utiliza um sintetizador de fala integrado com um processador de texto e com um acelerador de escrita, servindo de auxiliar ao ensino da Língua Portuguesa a crianças com paralisia cerebral. Esta ferramenta permite à criança uma escrita mais rápida e a possibilidade de ouvir ler o texto que acabou de escrever, permitindo a autocorrecção de erros de ortografia associados a sons que ela própria não consegue produzir.

A generalização do uso da fala como interface entre humanos e máquinas dependerá do aumento da naturalidade e sofisticação destes sistemas. Contudo, o estágio de desenvolvimento desta tecnologia para cada Língua depende do investimento efectuado no seu estudo. É natural que o Inglês continue como modelo de qualidade, que Línguas como o Português só atingirão daqui a alguns anos. Mas mesmo para a Língua Inglesa, ainda estamos longe das capacidades de comunicação do HAL 9000 imaginado por Arthur C. Clark em "2001: Uma Odisseia no Espaço".

* *Investigador do Instituto Nacional de Engenharia de Sistemas e Computadores*