Toc-Toc, Tic-Tac, Triiiimm!
Utilização de Som em Interfaces Multimodais

Carlos Laufer, Daniel Schwabe

Departamento de Informática – Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)
Caixa Postal 38097 – 22453-900 – Rio de Janeiro – RJ – Brasil

laufer@globo.com, dschwabe@inf.puc-rio.br

Abstract. *As information presented by the interfaces of computer and mobile device applications become more and more visually intensive, the visual channel is becoming increasingly overloaded and we become limited in our capacity of assimilating information. The sound has a significant role in our everyday life but has been slight explored in the way we interact with computers and mobile devices. This article presents a discussion on the necessity of integration of different sensory modes in multimodal interfaces, mainly the use of audio information, and address relevant concepts like auditory icons, earcons, attention, semiosis, abductive processes, anticipation, speech acts, etc.*

Resumo. *À medida que as informações apresentadas pelas interfaces das aplicações em computadores e dispositivos móveis se tornam, cada vez mais, visualmente intensivas, o canal visual fica sobrecarregado e nos tornamos limitados em nossa capacidade de assimilar informações. O áudio tem um papel significativo no nosso dia-a-dia, mas tem sido pouco explorado na forma como interagimos com o computador e com dispositivos móveis. Este artigo apresenta uma discussão sobre a necessidade da integração de diferentes modos sensoriais em interfaces multimodais, particularmente o uso de informações sonoras, e aborda conceitos relevantes como ícones auditivos, earcons, atenção, semiose, processos abdutivos, antecipação, atos de fala, etc.*

1. Introdução

À medida que as informações apresentadas pelas interfaces das aplicações executadas em computadores e dispositivos móveis se tornam, cada vez mais, visualmente intensivas, o canal visual fica sobrecarregado e nos tornamos limitados em nossa capacidade de assimilar informações. Existe, atualmente, um desenvolvimento de dispositivos móveis para utilização nas funções do dia-a-dia, com aplicações em diversas áreas como entretenimento, orientação espacial, negócios, etc. Em muitas situações, durante a manipulação desses dispositivos, não é possível manter-se um contato visual constante com esses aparelhos como, por exemplo, um mapa apresentado em um visor de navegação dentro de um automóvel.

O áudio tem um papel significativo no nosso cotidiano. Nós utilizamos as informações de áudio para perceber situações perigosas, atender telefonemas, diagnosticar problemas em nossos carros, atrair a atenção de pessoas, perceber a

presença de outros, etc. Esse valioso modo de percepção tem sido pouco explorado na forma como interagimos com o computador e com dispositivos móveis. A maioria das pessoas tem capacidade para monitorar simultaneamente um conjunto de informações sonoras, enquanto está realizando uma tarefa que exige atenção visual. Uma pessoa pode dirigir um carro, com o rádio ligado, enquanto conversa com outro passageiro do veículo. Mesmo concentrado na conversa, o motorista pode monitorar o que está ouvindo no rádio e, se for de seu interesse, interromper a conversa para comentar sobre uma música do seu agrado. Enquanto isso ocorre, o motorista pode estar ultrapassando outro veículo e, nesse processo, trocando de pista na rodovia. Um som repetido informa que a seta que indica mudança de direção está funcionando corretamente e, caso o carro tenha uma transmissão manual, o som do motor indicará quando é o momento adequado para se trocar a marcha. Além de tudo isso, o motorista pode perceber se o motor produz algum ruído estranho ou se uma ambulância se aproxima.

Uma pessoa pode extrair diversas informações a partir de um som recebido, podendo identificar diversas características. O som tem uma natureza temporal e, por se tratar de uma onda mecânica, ocorre a partir do movimento. O movimento gera som. Os objetos produzem sons característicos quando em movimento: o barulho de um motor, de um ar condicionado ligado, de um teclado de computador sendo acionado, dos passos de uma pessoa subindo uma escada. Os sons fornecem informações relacionadas à localização espacial de onde eles estão sendo gerados. A partir da audição dos sons dos passos de uma pessoa é possível localizá-la, informar se ela se aproxima ou se afasta, informar se está subindo ou descendo escadas. Também é possível extrair informações relacionadas à dimensão dos objetos—o ruído dos passos pode indicar o tamanho ou o peso de uma pessoa.

A exploração do uso de sons em interfaces encontra-se ainda numa fase bastante incipiente, se comparada com a maciça utilização dos recursos visuais. O aumento das informações a que uma pessoa é atualmente bombardeada diariamente, por meio da interação com um conjunto cada vez maior de dispositivos, muitos deles com capacidade móvel, torna urgente que, devido a capacidade cognitiva limitada dos seres humanos de resposta aos estímulos, todos os modos sensoriais sejam explorados ao máximo, com o objetivo de auxiliar, da melhor forma possível, o receptor dessas informações. Além disso, muitas das vezes, os usuários se encontram em movimento e não têm uma possibilidade de contato visual constante com os dispositivos [Brewster and Walker 2000].

O desafio de grandes volumes de dados heterogêneos tem muitas facetas. Sua armazenagem, arquitetura, mecanismos de recuperação e apresentação, em diversos dispositivos, ocupam normalmente o foco de pesquisas e discussões. Porém, igualmente importante é atentar para como o ser humano, usuário e destinatário final de todos os benefícios trazidos por esse tipo de tecnologia, lidaria com a informação produzida e movimentada por sistemas dessa natureza.

Este trabalho se destina a discutir uma faceta específica desse segundo tipo de questão, relacionada ao desafio: possibilidades de representação de grandes volumes de dados heterogêneos em mídias alternativas. Parte importante da discussão é a necessidade de as representações e mídias não competirem pela atenção perceptiva ou

cognitiva do usuário, mas, ao contrário, integrarem-se adequadamente para facilitar o acesso e aproveitamento desses grandes volumes de dados [Brown et al. 1989].

A proposta deste trabalho é estabelecer um referencial teórico que permita projetar e implementar, de forma sistemática, mecanismos que explorem o desenvolvimento de estudos e protótipos que averigúem o potencial das habilidades auditivas dos seres humanos. Por meio desses mecanismos, o áudio pode melhorar a qualidade da interação humana com sistemas complexos, seguindo uma tendência atual de construção de interfaces de usuário multimodais. As indicações sonoras podem vir a ter um papel importante no aumento da capacidade de absorção de informações por parte dos usuários. Nosso foco trata especificamente de um aspecto do som que é bastante negligenciado: incrementar a utilização de áudio com informação sonora não-verbal, para comunicar informação aos usuários de computadores e dispositivos móveis, visando aproximar o uso do áudio do patamar utilizado no cotidiano das pessoas.

Nas seções seguintes examinaremos as diversas facetas e fatores que são relevantes ao uso de áudio (e, em muitos casos, de informação multimídia em geral) em interfaces humano-computador.

2. Limitação da Capacidade Cognitiva e Atenção

Diversos estudos na área da psicologia têm seu foco em cognição e atenção. Os psicólogos cognicistas estão interessados nos eventos que ocorrem entre a apresentação de um estímulo e o desempenho em uma resposta respectiva. A psicologia cognitiva contemporânea traça em muitas ocasiões uma analogia entre o homem e o processamento de informações que ocorre dentro de um computador. Existem diversos processos cognitivos, entre eles: atenção, reconhecimento de padrões, memória de curta duração, memória de longa duração, raciocínio e processamento de linguagem.

A atenção é nossa habilidade em focar certos aspectos da experiência cotidiana e imediata, enquanto ignoramos outros aspectos. Ela é crucial para isolarmos alguma coisa que desejamos perceber, perante os diversos estímulos aos quais estamos submetidos constantemente. Segundo James Williams, em seu livro Princípios de Psicologia [Williams 1890], "Todos sabem o que é atenção. É a posse pela mente, de forma clara e vívida, de uma dentre diversas possibilidades de objetos ou raciocínios. O foco e a concentração são a essência da consciência. Isso implica no descarte de algumas coisas, para lidar de maneira efetiva com outras e é uma condição que tem um oposto real, no confuso, aturdido, desatento estado que, em francês, é chamado de distração."

Os seres humanos têm uma capacidade bastante limitada para processar informações sensoriais. A teoria do gargalo (*bottleneck*), desenvolvida por Welford [Pashler 1995], investiga as dificuldades humanas na realização de tarefas simultâneas. Pode uma pessoa realizar simultaneamente duas tarefas com a mesma qualidade e desempenho? Que fatores podem afetar a habilidade de uma pessoa em realizar duas tarefas ao mesmo tempo? Como o sistema atencional de uma pessoa controla o desempenho de duas tarefas?

A atenção pode ser definida como a habilidade para selecionar parte da informação que é recebida a partir de um ou mais estímulos, para que haja um processamento mais aprofundado. Portanto, a atenção se refere a sistemas cognitivos

que nos permitem selecionar e processar uma informação específica, enquanto outras são ignoradas por serem julgadas de menor relevância ou importância. A atenção pode ser classificada em dois tipos principais: atenção seletiva ou focalizada—o foco está em apenas uma parte do ambiente—e atenção dividida, onde não existe um único foco de atenção—a atenção encontra-se espalhada por dois ou mais estímulos [Kahneman 1973] [Becklen 1983]. Além disso, é possível distinguir a atenção entre voluntária e involuntária. A atenção voluntária é movida por um objetivo deliberado da pessoa, enquanto a atenção involuntária ocorre quando alguma informação do ambiente captura sua atenção como, por exemplo, um barulho repentino ou o som de nosso nome sendo falado em uma outra conversa. Os comportamentos humanos emergem a partir da interação dos objetivos que uma pessoa possui e de estímulos que ocorrem vindos do ambiente [Pashler et al. 2001].

A partir de meados do século XX, diversos modelos foram criados para explicar o funcionamento do sistema sensorial e da limitação da sua capacidade de processamento [Driver 2001]. Um artigo pioneiro de Cherry (1953) aborda o "efeito coquetel" (*cocktail party effect*), que investiga como uma pessoa, em um ambiente repleto de conversações paralelas e simultâneas, pode selecionar uma determinada conversa, em detrimento das outras. Além disso, mesmo estando em uma determinada conversa, uma pessoa pode ter sua atenção direcionada a outra conversa, caso ouça alguma informação que lhe é importante ou familiar como, por exemplo, seu próprio nome. Em 1958, Donald Broadbend lança a teoria do filtro de atenção, segundo a qual, o sistema sensorial de uma pessoa receberia dois estímulos diferentes de forma paralela. Como o sistema tem uma capacidade limitada, ele permitiria que apenas um dos estímulos, a partir de suas propriedades físicas, passasse por um filtro, sendo que o outro estímulo ficaria armazenado em um *buffer*, para processamento posterior. Em seguida, Treisman and Gelade (1960) definem a teoria do filtro atenuador, onde as mensagens não-atendidas não são totalmente descartadas, mas são apenas atenuadas, sendo todas elas processadas pelo sistema central. Em um artigo publicado em 1963, J. Anthony Deutsch e Diana Deutsch definem a teoria da seleção posterior, onde todos os estímulos percebidos são processados integralmente, sendo que a ação é determinada com base na relevância de cada estímulo para a situação.

Diversos fatores influenciam a atenção de uma pessoa: aspectos do estímulo em si, aspectos próprios da pessoa e as interações entre estímulos específicos e as experiências e interesses da pessoa. Aspectos relacionados a um estímulo são um dos componentes que podem atrair a atenção. Por exemplo, a intensidade e a duração de um som podem influenciar a percepção desse som. Além disso, variações e repetições também podem chamar a atenção.

O estado interno de uma pessoa pode calibrar suas percepções. Uma pessoa quando está com fome fica muito mais sensível a perceber objetos comestíveis ou até a se perturbar com o barulho feito por uma pessoa comendo. Para uma pessoa cuidando de uma criança pequena, qualquer barulho diferente soa como um alarme. Uma pessoa que não tenha relação alguma com aquela criança terá uma calibragem sensorial completamente diferente.

As posturas e as ideias de uma pessoa são outros dois fatores que influenciam na determinação de quais aspectos do ambiente essa pessoa irá notar. Existe um ajuste que

frequentemente não é nem mesmo consciente. A experiência passada prepara a pessoa para responder aos estímulos de uma forma particular. Sua experiência passada a leva a esperar detemirnadas coisas, a antecipar determinados estímulos. Por exemplo, você percebeu que a palavra "determinadas" estava grafada incorretamente na frase anterior? Essa característica pode ser benéfica ou não. No caso da leitura de texto, pode levar a uma rapidez maior de leitura. Se for um texto relacionado a entretenimento, não traria maiores problemas, porém, se for um texto legal, como um contrato de locação de um imóvel, é aconselhável uma atenção, uma concentração maior, pois a palavra é o texto da lei.

A situação, o cenário, em que uma pessoa se encontra, pode influenciar sua percepção dos fatos. Uma situação de pressão pode alterar a calibragem sensorial de uma pessoa. Durante uma partida de futebol, quando um determinado atacante de um dos dois times é derrubado dentro da grande área, cada torcedor tem uma percepção bastante parcial quanto à ocorrência ou não da falta, dependendo do time de sua preferência.

Existem aspectos dos estímulos que se combinam com a experiência anterior da pessoa, para determinar o que irá atrair sua atenção. Quando alguém, que você não conhece, se apresenta dizendo como se chama, caso seja um nome de uso comum como, por exemplo, "Carlos", é bem possível que você entenda o nome com facilidade, sem necessidade de muita atenção. Porém, se a pessoa se apresenta como "Laufer", existe uma boa chance de você não entender imediatamente esse nome, sendo que em muitos casos, você pedirá para a pessoa repetir o nome, com você adotando uma postura de atenção redobrada, para poder perceber a fala da outra pessoa. Duas pessoas que convivem, muitas vezes entendem as frases uma da outra, antes mesmo que as frases sejam completadas.

Essa mescla de fatores também influencia o foco de sua atenção em um ambiente. Por exemplo, um trabalhador que sabe que o fim do seu turno de trabalho é sinalizado pelo som de uma sirene, pode ficar mais sensível a esse som quando se aproxima a hora de ir para casa. Um torcedor de futebol que tem seu time ganhando um jogo por um placar bem apertado, numa partida final da Copa do Mundo, ao se aproximar o fim do jogo, fica bastante sensível ao apito do árbitro que encerra a partida.

3. Áudio Verbal, Ícones Auditivos, *Earcons* e Sonificação

Ouvir o tom de uma música é um exemplo de audição musical. Entretanto, nós frequentemente ouvimos eventos ao invés de sons. Ouvir o barulho de aviões, de água, de pássaros e de passos são exemplos de audições cotidianas [Gaver 1988]. Esse é um tipo de experiência diferente daquela descrita pela psicoacústica tradicional. Ao invés de estar relacionado a nossa habilidade de perceber os atributos dos sons em si—frequência, amplitude, etc.—, a audição cotidiana está relacionada aos atributos dos eventos que ocorrem no mundo: a velocidade de um carro que passa, a força de uma porta batendo, uma pessoa pesada subindo ou descendo os degraus de uma escada, entre outros.

Historicamente, os estudos de acústica e psicoacústica foram guiados com uma preocupação maior com o entendimento da música e dos sons produzidos por instrumentos musicais. Os estudos da estrutura harmônica dos sons musicais nas

disciplinas direcionadas ao áudio são ligados aos sons musicais e ao entendimento da audição musical. Mas seria essa a melhor forma de descrever eventos sonoros que escutamos durante o dia-a-dia? A teoria acústica e psicoacústica têm sua ênfase em dimensões relacionadas à percepção e à parte física, que são mais adequadas para a descrição de música. Os sons musicais parecem prover pouca informação a respeito de suas fontes, enquanto os sons cotidianos frequentemente fornecem uma grande quantidade de informação sobre eles.

Os estudos relacionados à utilização de sons em interfaces classificam o áudio em três categorias: áudio verbal, ícones auditivos e *earcons* [Gaver 1989]. O áudio verbal está ligado à utilização da fala propriamente dita. Os ícones auditivos se relacionam à utilização de sons do cotidiano nas interfaces, em metáforas e analogias do mundo real. Blattner et al. (1989) definem *earcons* como "mensagens de áudio não-verbais que são utilizadas em interfaces de computador/usuário para prover informações ao usuário sobre algum objeto, operação ou interação computacional". Diferentemente de um ícone auditivo, não existe um elo intuitivo entre um *earcon* e aquilo que ele representa. Em geral, os *earcons* utilizam um enfoque mais musical do que os ícones auditivos. Os *earcons* são sons associados às características físicas do som e não aos eventos do cotidiano. Por exemplo, quando um usuário esvazia a lata de lixo do seu *desktop*, um aviso sonoro poderia ser apresentado de três maneiras: uma voz pré-gravada dizendo "Os arquivos existentes na sua lata de lixo foram apagados"; o áudio de uma lata de lixo sendo esvaziada em um caminhão de recolhimento de lixo; um simples bip.

A sonificação é o processo de utilização de áudio não-verbal como forma de disponibilizar informações [Flowers et al. 2005] [Walker and Nees 2009]. Uma das primeiras aplicações de sucesso a utilizar sonificação foi o Contador Geiger, um dispositivo para a medição de radiação, onde a frequência dos clics apresentada pelo dispositivo é diretamente proporcional ao nível de radiação no ambiente. Devido a características da percepção auditiva—resolução temporal, espacial, etc.—a sonificação se aplica com sucesso em situações que requerem uma constante monitoração de informação, por exemplo, as funções vitais do corpo, durante operações cirúrgicas.

As primeiras definições de sonificação caracterizavam essa técnica, basicamente, como um mapeamento de uma massa de dados segundo uma perspectiva sonora, como forma análoga à perspectiva visual. Em Barras (2005) é apresentado um *framework* para a representação de dados científicos de forma sonora.

Hermann (2008) define que uma técnica de geração de sons pode ser chamada de sonificação, se essa técnica utiliza dados como entrada e gera sinais sonoros como resposta, de acordo com as seguintes premissas: esses sons refletem propriedades ou relações objetivas dos dados de entrada; a transformação é sistemática (existe uma definição precisa de como os dados fazem o som ser alterado); a sonificação pode ser reproduzida (um mesmo dado para as mesmas interações tem uma sonificação estruturalmente idêntica). Segundo essa definição, Hermann inclui a utilização de ícones auditivos e *earcons* como possibilidades de apresentação de dados e, portanto, como possibilidades de sonificação.

4. Interfaces Mutimodais

O mundo a nosso redor nos fornece um fluxo contínuo de estímulos, captados por todos os nossos sentidos. Objetos e eventos podem ser vistos, ouvidos, cheirados, tocados, degustados e, à medida que nos movemos e interagimos com pessoas, locais e objetos em nosso ambiente, produzimos mudanças constantes nas nossas atividades. Estudos na área da neuroanatomia e da neurofisiologia indicam que a junção de estímulos temporais e espaciais, a partir de modalidades sensoriais diferentes, pode levar a uma resposta neural que é maior do que a soma das respostas neurais aos componentes unimodais da estimulação, quando considerados separadamente. Ou seja, a atividade de um neurônio exposto a uma estimulação multissensorial, por exemplo, uma estimulação visual e auditiva simultâneas, difere de forma significativa da atividade da mesma célula quando exposta à estimulação individual, para qualquer uma das duas modalidades [Bahrick and Lickliter 2002].

As interfaces multimodais envolvem a utilização de diferentes modalidades humanas na interação entre o usuário e um computador ou dispositivo. Diversas iniciativas pesquisam a utilização de interfaces multimodais em dispositivos e sistemas. Muitas dessas iniciativas se aplicam ao incremento da comunicação com pessoas portadoras de deficiências dos sentidos. Um dos trabalhos pioneiros nesta área foi uma aplicação chamada Soundtrack [Edwards 1989]—um editor de textos com a apresentação de informações utilizando uma interface sonora. Murphy et al. (2007) apresentam um *plug-in* para navegadores de Internet, que utiliza sons como forma de comunicação com pessoas com deficiências visuais.

Com o crescimento de dispositivos que apresentam possibilidades de comunicação háptica, aumentam os trabalhos que combinam elementos visuais, sonoros e hápticos, na comunicação estabelecida entre os sistemas e os usuários. Um telefone celular já combina características hápticas e sonoras, quando sinaliza a chegada de uma nova chamada telefônica, também, por vibração. McGee et al. (2000) apresentam um estudo de como incrementar a sensação de texturas, utilizando um dispositivo que permite que o usuário receba informações hápticas (*Phantom Force Feedback*) em conjunto com elementos sonoros.

Wearable Computers é uma linha de pesquisa relativa a dispositivos que utilizam periféricos que podem ser "vestidos" pelo usuário, incrementando, assim, o modo como as informações podem ser transmitidas ao sistema: fones de ouvido com capacidade de detectar os movimentos da cabeça do usuário, possibilitando assim que o usuário possa se comunicar com o sistema a partir de gestos com a cabeça; agendas eletrônicas com detecção de movimento e posicionamento espacial, permitindo assim que movimentos, como o chacoalhar do dispositivo, possam ser interpretados pelo sistema; etc. [Brewster 2005].

5. Semiose e Processos Abdutivos

Diariamente, estamos imersos em uma miríade de sons que nos trazem informações das mais diversas. Muitos desses sons são gerados de forma não-intencional como, por exemplo, os sons da vassoura de um gari varrendo o chão da rua, de um carro passando, do motor de um ônibus, entre outros. Existe uma outra categoria, que engloba sons gerados de forma intencional como, por exemplo, a sirene de uma ambulância, o badalar

dos sinos de uma igreja, o toque da corneta do quartel de bombeiros [Walker and Nees 2009]. Tanto os sons gerados de forma não-intencional como os gerados de forma intencional carregam informações que são entendidas pelos humanos, dentro de um sistema de significação.

Um signo é alguma coisa que representa algo para alguém. A teoria geral dos signos procura explicar o significado do significado. Duas das principais linhas de pesquisa relacionadas à teoria dos signos tiveram seu início no começo dos anos 1900. Saussure (1910) define uma linha, a semiologia, ligada à interpretação dos signos de uma linguagem. Ele define um modelo diádico do signo, composto por um significado e um significante. Peirce define uma linha ligada à lógica, denominada semiótica [Santaella 2006]. Ele define um modelo triádico para o signo, composto por um objeto, uma representação e um interpretante. Peirce define que o processo de significação de um signo, a semiose, é um processo infinito. O interpretante de um determinado signo é também um signo que, por sua vez, tem uma relação triádica com o objeto e um novo interpretante, e assim por diante.

Segundo Peirce existem três modos de raciocínio: a dedução, a indução e a abdução. A abdução é um processo no qual uma pessoa, ao se deparar com um fato, estabelece uma hipótese para a sua significação e, a partir da exclusão das possibilidades de falha dessa hipótese, conclui que a mesma é verdadeira. Caso a primeira hipótese falhe, uma nova hipótese é imediatamente estabelecida e todo o processo se repete, num mecanismo, de alguma forma, análogo à semiose. Por exemplo, ao retornar do trabalho, à noite, uma pessoa, antes de abrir a porta de casa, percebe as luzes acesas dentro da sua residência. Ela pode formular, por exemplo, a seguinte hipótese: “existe alguém em casa”. Ao entrar em casa, verifica que não existe ninguém e reformula sua hipótese, considerando que “alguém deve ter deixado as luzes acesas, ao sair de casa”. Porém, se, ao entrar no quarto, percebe que todas as gavetas foram reviradas, uma nova hipótese plausível é a de que “houve uma invasão da residência”.

Qualquer informação percebida por um dos sentidos humanos—visão, audição, tato, olfato e paladar—pode auxiliar no processo abdutivo cotidiano de uma pessoa. Quando um som alcança um humano, ele pode disparar uma ação por parte do humano, dependendo do contexto em que ele se encontra. Se estou dirigindo meu carro e escuto o som de uma sirene de ambulância, eu entendo que devo manobrar meu veículo de modo a dar passagem à ambulância. Se, durante a tarde, estou trabalhando no computador de casa e ouço o badalar dos sinos da igreja do bairro, entendo que são seis horas da tarde e que está na hora de eu ir dar minha volta de bicicleta na ciclovia da orla da praia. Se estou tomando o café da manhã em um hotel, pela primeira vez, e estou procurando onde estão os talheres, ao escutar o som característico de uma pessoa pegando talheres, percebo de onde se origina o som e, dessa forma, localizo o que procuro.

Seres humanos agem, em muitas situações, a partir de uma antecipação [Nadin 2003]. Em muitos casos, mesmo quando não percebemos, antecipamos uma situação que gera algum tipo de efeito. Por exemplo, quando estamos em um elevador, o nosso corpo se prepara para o movimento de subida ou descida do elevador. Na situação de anteciparmos que o elevador irá subir e, ao invés disso, o elevador descer, sentimos um certo desconforto físico, pois nosso corpo foi “antecipado” para um movimento de subida. As pessoas antecipam possibilidades, estabelecem hipóteses. Se, por exemplo,

em uma corrida de carros da Fórmula 1, surgem indícios de chuva—nuvens densas e escuras, vento forte, etc.—, alguma equipe pode decidir se preparar para a possibilidade de chuva e colocar os pneus adequados a essa possibilidade. A decisão da escolha da possibilidade, da antecipação, pode se fundamentar em indícios baseados em, previsões, estudos probabilísticos, intuição, experiência, etc. Porém, uma vez escolhida a possibilidade, as ações tomadas se adequarão à antecipação definida.

6. Atos de Fala

Uma conversação estabelecida entre duas pessoas por meio de frases construídas em uma determinada língua é constituída por atos de fala [Austin 1962]. Um ato de fala contêm três tipos de atos que são classificados como: atos locucionários, atos ilocucionários e atos perlocucionários. Os atos locucionários são compostos da articulação da frase e da proposição que essa contém. O ato ilocucionário está ligado à intenção que se pretende dar ao ato locucionário. O ato perlocucionário está ligado a alguma ação que o ouvinte de um ato de fala possa vir a tomar, a partir de seu recebimento. Os atos ilocucionários podem ser classificados segundo a seguinte taxonomia: assertivas, diretivas, comissivas, expressivas e declarativas [Searle 1969]. A força ilocucionária de um ato de fala pode ser transmitida por diversos meios: um verbo ilocucionário, a entonação utilizada, o contexto no qual se insere o ato de fala, etc.

Cada som não-verbal poderia ser, de alguma forma, traduzido para um ou mais atos de fala. Por exemplo, o som emitido quando passamos com algum objeto metálico pelo detector de metais de um aeroporto poderia ser traduzido pela expressão: “Um objeto metálico foi detectado”. Todo ato de fala tem uma intenção. Considerando a taxonomia dos atos ilocucionários, a expressão “Um objeto metálico foi detectado” seria uma assertiva. Além disso, o bip do detector de metais embute um outro ato ilocucionário, que poderia ser traduzido pela expressão: “Por favor, retorne e verifique se você esqueceu de retirar algum objeto metálico antes de passar pelo detector”. Considerando a taxonomia dos atos ilocucionários, essa expressão seria uma diretiva. Ao invés de emitir o som, que é identificado pelos operadores do aeroporto como sinal de que algo foi detectado, o dispositivo detector de metais poderia ter uma voz gravada dizendo: “Um objeto metálico foi detectado. Por favor, retorne e verifique se você esqueceu de retirar algum objeto metálico antes de passar pelo detector”. Porém, o sinal sonoro é mais conciso e de melhor identificação e percepção, uma vez que esteja no sistema de significação dos ouvintes.

Vamos supor, agora, um exemplo inverso ao anterior. Um motorista tem um veículo com dispositivo para auxílio à navegação, que utiliza GPS. O motorista deseja ir a um local e informa o endereço de destino para o dispositivo. Com a navegação em curso, no instante em que o dispositivo identifica a necessidade de se virar à esquerda, no próximo cruzamento, ele comunica ao motorista, por meio de uma voz gravada: “Vire à esquerda no próximo cruzamento”. Considerando que os automóveis possuem, usualmente, sistemas estéreos de som, essa frase poderia ser substituída por um sinal sonoro característico, que seria emitido do lado esquerdo do sistema de som do motorista. Com algum tempo de treinamento, o motorista poderia ser informado dessas manobras de forma não-verbal. Neste caso o sinal sonoro funcionaria com um ato de fala.

7. Contexto

Os sons não-intencionais estabelecem no usuário uma consciência a respeito do ambiente no qual ele está inserido. Em geral, o contexto define um grau mais acentuado de atenção, uma acuidade de percepção, em relação a determinados sons. Com o crescimento do número de plataformas de computação portáteis, de dispositivos de comunicação móveis, e da combinação das plataformas com os dispositivos, fica cada vez mais presente o conceito de sistemas que consideram a característica nômade dos usuários [Kleirock 1996]. Esses sistemas devem considerar a possibilidade dos usuários estarem conectados a partir de diversos pontos físicos: sua casa, escritório, automóvel, um vagão do metrô, etc. O contexto de cada um desses ambientes é diferente e deve ser levado em consideração pelos sistemas.

A computação sensível ao contexto é uma área de pesquisa relacionada a sistemas que coletam informações contextuais para o auxílio mais efetivo e eficiente ao usuário. Abowd et al. (1999) definem contexto como "qualquer informação que possa ser utilizada para caracterizar uma situação de uma entidade. Uma entidade é uma pessoa, lugar ou objeto que é considerado relevante para a interação entre o usuário e a aplicação". A locação, a identidade, a data e hora e a atividade são os tipos primários de contexto, para fins de determinação de uma entidade particular. Eles respondem a questões tais como, quem, o que, quando e onde, e também servem como índices para outras fontes de informação contextual. A partir da identidade de uma pessoa, poderia se ter acesso a seu número de telefone, endereço, etc. Dey (2001) define que um sistema é sensível a contexto se ele utiliza o contexto para prover informações e/ou serviços relevantes para o usuário, onde a relevância depende da tarefa em que usuário está engajado.

Want et al. (1992) desenvolveram um trabalho pioneiro nessa área—uma rede de sensores instalados em um ambiente de trabalho podia captar sinais eletrônicos emitidos por crachás utilizados pelos funcionários. Dessa forma, era possível, por exemplo, redirecionar uma chamada telefônica para um telefone próximo ao local onde o funcionário se encontrava num determinado instante, ou saber se o funcionário estava presente no escritório. O Hippie foi um outro exemplo de sistema de informação nômade, desenvolvido para fornecer informações sobre exibições de arte em um ambiente cultural—o usuário nômade tem suporte de informações adaptativas durante todo o processo de visita a um museu [Oppermann 2001]. O contexto considerado pelo sistema leva em conta informações colhidas junto ao usuário (preferências quanto a roteiros predeterminados de visitas, etc.) e a localização física das obras, a partir de uma rede de sensores espalhados pelo ambiente. Dessa forma, é possível fornecer informações detalhadas sobre as obras, à medida que o usuário percorre o espaço físico do museu.

Os sons podem estar presentes em um ambiente, mas podem não ser percebidos por uma pessoa, se o espaço de sons perceptíveis que condiz com o contexto corrente dessa pessoa não contempla aquele tipo de som. No exemplo do café da manhã no hotel, o som de talheres sendo manuseados está presente com bastante frequência no ambiente, mas, em geral, passa despercebido. Fica como um som de fundo misturado aos demais sons que formam o burburinho do ambiente. No caso de se necessitar dos talheres, esse

som adquire um potencial maior de ser percebido, pois, no sistema de significação da pessoa, ele está relacionado ao objeto sendo procurado.

O contexto é um fator fundamental no processo de significação. Podemos considerar que existem dois tipos de contexto para uma determinada situação cotidiana envolvendo um usuário: um contexto semântico e um contexto sensorial. O contexto semântico está relacionado à situação, ao cenário, envolvendo o usuário e o contexto sensorial está relacionado aos níveis de percepção do usuário. No caso do exemplo dos talheres, o contexto semântico estaria relacionado à procura por talheres e o contexto sensorial seria o aumento do nível de percepção auditiva e visual em relação aos talheres. Considerando a existência de uma acuidade de percepção relacionada a cada um dos sentidos humanos, podemos imaginar que cada uma dessas acuidades se ajusta de acordo com um determinado contexto semântico.

O contexto sensorial seria resultante da composição de todas essas acuidades, uma calibragem de cada um dos sentidos que estaria influenciada por diversos fatores. Podemos imaginar uma espécie de nível mínimo de percepção de cada um dos sentidos, uma espécie de limiar, abaixo do qual eventos relacionados àquele determinado sentido não seriam percebidos, ou não teriam uma atenção maior. Em contrapartida, determinados contextos semânticos definiriam contextos sensoriais que estariam calibrados para uma solução mais eficaz e eficiente da situação associada ao contexto semântico. O contexto sensorial poderia ter uma contrapartida na própria interface de uma aplicação ou dispositivo. A partir de um determinado contexto semântico, em um determinado momento da interação do usuário, um contexto sensorial da interface definiria um grau de intensidade maior, ou menor, para as informações disponibilizadas pela interface, numa espécie de negativo do contexto sensorial do usuário. Considerando o exemplo dos talheres, seria como se na situação em que o usuário procura os talheres, o som dos talheres ganhasse maior intensidade.

8. Metáforas Sonoras Espaciais

Uma das informações que podem ser identificadas quando uma pessoa escuta um som é a posição dessa fonte sonora. Ao ouvir a sirene de uma ambulância, uma pessoa dirigindo seu carro é capaz de perceber, por exemplo, se a ambulância se aproxima pela mesma via onde se encontra seu veículo ou por uma via transversal. Uma pessoa andando pela calçada, ao ouvir alguém gritando seu nome, gira automaticamente seu corpo na direção de onde o som se origina.

Essa característica é utilizada em diferentes sistemas de reprodução de música. Os primeiros sistemas de som utilizavam apenas um canal de gravação, o que gerava um som chamado de mono. Nesse tipo de sistema não é possível se ter uma ideia do posicionamento espacial da fonte sonora. O primeiro sistema a utilizar de forma mais precisa o conceito posicional de som foi o sistema estéreo, que permite o posicionamento de um determinado som numa dimensão horizontal. São dois canais independentes de áudio reproduzidos por sistemas de som dispostos simetricamente. É possível, dessa forma, se ter a sensação de um som posicionado mais à esquerda ou mais à direita. A evolução dos sistemas estéreos se deu pela criação dos sistemas que utilizam mais canais para a reprodução do som, por meio de sistemas de autofalantes distribuídos na frente, atrás, acima, abaixo e ao lado dos ouvintes. O som dos filmes, jogos de

computador, etc., utilizam tecnologias como, por exemplo, o sistema de som 5.1, que utiliza 6 canais independentes para a reprodução do som. A forma de gravação de som que mais aproxima a característica 3D de um som, ou seja, seu posicionamento no espaço, é a gravação binaural. Sons gravados de acordo com esse conceito permitem, com o uso de fones de ouvido, que se tenha uma noção espacial bastante sofisticada, quando se ouve a reprodução do som.

A maioria dos sons que são utilizados nas interfaces dos dispositivos não faz uso dessas novas tecnologias e, portanto, desperdiça a capacidade plena de posicionamento espacial de um som reproduzido. A utilização da característica 3D do som poderia incrementar as informações fornecidas nas interfaces de aplicações e dispositivos. Em uma sessão de msn, skype, etc., o posicionamento do áudio emitido pelos participantes da sessão poderia estar de acordo com uma distribuição espacial definida pelo usuário. Por exemplo, o som emitido pelo principal assessor poderia ser posicionado como se ele estivesse localizado à direita, bem próximo ao usuário, enquanto o som emitido por um funcionário de menor escalão poderia ser posicionado como se ele estivesse localizado numa posição bem mais afastada.

Essa distribuição espacial também poderia estar associada à chegada de e-mails. Por exemplo, um e-mail do principal assessor de um usuário poderia emitir um som localizado bem próximo e à direita do usuário, enquanto um e-mail de um amigo informal poderia emitir um som localizado bem mais distante e ao fundo. Um refinamento dessa utilização seria possibilitar a configuração espacial sonora de acordo com um contexto relacionado a um cenário específico, criando, assim, esquemas particulares. No exemplo da interface sonora para a chegada de e-mails, poderíamos ter dois esquemas: empresa e residência. Quando o usuário, durante a semana, estivesse no seu ambiente de trabalho, vigoraria o esquema empresa, descrito acima. Durante o fim de semana, vigoraria o esquema residência, onde a configuração espacial sonora poderia ser a do amigo numa posição próxima, enquanto o principal assessor estaria numa posição bem distante, ou mesmo, sem som algum.

9. Conclusão

Neste artigo examinamos diversos aspectos que têm relevância e devem ser considerados ao projetar mecanismos para o uso de áudio nas interfaces humano-computador. A seguir, tecemos algumas considerações que ilustram como esses aspectos poderiam ser usados para identificar de forma mais precisa as diversas dimensões de projeto a serem abordadas.

A partir das discussões prévias, podemos (re)enunciar o problema de uso de áudio em interfaces como sendo o de como apoiar, com o uso de sons, o processo abdutivo e as antecipações que um usuário realiza durante a execução de tarefas auxiliadas por dispositivos com interfaces multimodais. É impossível limitar o espaço de hipóteses que um usuário pode criar, como forma de realizar uma ação. Não é possível prever com exatidão o processo abdutivo de um determinado usuário. Porém, podemos, levando em conta fatores culturais, experiência de uso, treinamento, etc., considerar a possibilidade de colocar indícios, pistas, que possam restringir ou, de alguma forma, guiar, orientar, esse espaço.

O processo de atribuição de sons em uma interface poderia levar em consideração as características dos atos de fala em conjunto com a característica humana do processo abdutivo de raciocínio e as antecipações. Um som utilizado em uma comunicação com um usuário pode disparar uma reação ou, então, auxiliar o processo de antecipação de uma ação futura. No caso das antecipações, os sons poderiam ser utilizados de maneira a orientar o processo abdutivo do usuário, diminuindo o espaço de hipóteses formuladas. Considerando o exemplo do dispositivo navegador GPS, poderíamos, a partir de um sistema de significação de sons, gerar um mapa de sons que pudesse ser compreendido pelo motorista e o auxiliasse no processo de navegação. O dispositivo navegador poderia ter indicações de retas prolongadas, curvas que se aproximam, declives acentuados, etc., que poderiam complementar sinalizações visuais existentes nas rodovias e ser eventualmente úteis em situações de pouca visibilidade ou de sinalizações rodoviárias visuais insuficientes.

A consideração do contexto, visto pelo lado do sistema (em termos da percepção de características do ambiente e do usuário) e visto pelo lado usuário (em termos de suas preferências, relações de trabalho, relações pessoais, etc.) pode introduzir novas formas de interação entre usuários e sistemas e entre usuários entre si [Winograd 2001] [Greenberg 2001].

As indicações sonoras e visuais podem ser utilizadas de forma complementar para o provimento de informações. A tecnologia para a produção de sons está bastante desenvolvida e existe um grande conhecimento acerca da natureza e da percepção dos sons. O desafio que se estabelece é o da integração desse conhecimento e dessa tecnologia para a definição de interfaces homem-máquina multimodais que sejam úteis e utilizáveis.

Agradecimentos

Carlos Laufer é beneficiário de auxílio financeiro da CAPES – Brasil, Programa Nacional de Pós-Doutorado (PNPD), projeto PNPD0086088. Daniel Schwabe tem suporte parcial de bolsa de pesquisa do CNPq.

Referências

Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M. and Steggles, P. (1999). "Towards a Better Understanding of Context and Context-Awareness**",** Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing, Karlsruhe, Germany, Lecture Notes In Computer Science, vol. 1707, p. 304-307, Springer Berlin, Heidelberg.

Austin, J.L. (1975). "How to do Things with Words", Cambridge, MA, Harvard University Press.

Bahrick, L. E. and Lickliter R. (2002). "Intersensory Redundancy Guides Early Perceptual and Cognitive Development", Advances in Child Development and Behavior, vol. 30. p. 153-187, Elsevier B.V., Academic Press.

Barras, S. (2005). "A perceptual framework for the auditory display of scientific data", Transactions on Applied Perception (TAP), vol. 2, n° 4, p. 389-402, ACM, New York, USA.

Becklen, R. and Cervone, D. (1983). “Selective Looking and the Noticing of Unexpected Events”, Memory & Cognition, vol. 11, p. 601-608.

Blattner, M. M., Sumikawa, D. A. and Greenberg, R. M. (1989). “Earcons and Icons: Their Structure and Common Design Principles”, Human-Computer Interaction vol. 4, n° 1, p. 11-44, L. Erlbaum Associates Inc., Hillsdale, NJ, USA.

Brewster, S.A. and Walker, V.A. (2000). “Non-Visual Interfaces for Wearable Computers”, IEE Workshop on Wearable Computing (00/145), IEE Press.

Brewster, S.A. (2005). “Multimodal Interaction and Proactive Computing”, In British Council Workshop on Proactive Computing, Nizhny Novgorod, Russia.

Brown, M., Newsome, S. and Glinert, E. (1989). “An Experiment into the Use of Auditory Cues to Reduce Visual Workload”, Proceedings of the CHI '89 Conference on Human Factors in Computer Systems, New York, ACM, p. 339-346.

Cherry, E. C. (1953). “Some Experiments on the Recognition of Speech with One and with Two Ears”, The Journal of the Acoustical Society of America, vol. 25, n° 5, p. 975-979, September.

Dey, A. K. (2001). “Understanding and Using Context”, Personal and Ubiquitous Computing, vol. 5, n° 1, p. 4-7, Springer-Verlag London Ltd., February.

Driver, J. A. (2001). “Selective Review of Selective Attention Research from the Past Century”, British Journal of Psychology, vol. 92, England, p. 53-78, The British Psychological Society.

Edwards, A. D. N. (1989). “Soundtrack: an auditory interface for blind users”, Human-Computer Interaction , vol. 4, n° 1, L. Erlbaum Associates Inc., Hillsdale, NJ, p. 45-66.

Flowers, J. H., Buhman, D. C. and Turnage, K. D. (2005). “Data sonification from the desktop: Should sound be part of standard data analysis software?”, Transactions on Applied Perception (TAP), vol. 2, n° 4, p. 467-472, ACM, New York, USA.

Gaver, W. W. (1988). “Everyday Listening and Auditory Icons”, Doctoral Dissertation, University of California, San Diego.

Gaver, W.W. (1989). “The SonicFinder: An Interface that Uses Auditory Icons”, Human-Computer Interaction, vol. 4, n° 1, p. 67-94, L. Erlbaum Associates Inc., Hillsdale, NJ, USA.

Greenberg, S. (2001). “Context as a Dynamic Construct”, Human-Computer Interaction, vol. 16, n° 2, p. 257-268, L. Erlbaum Associates Inc., Hillsdale, NJ, USA.

Hermann, T. (2008). “Taxonomy and Definitions for Sonification and Auditory Display”, Proceedings of the 14th International Conference on Auditory Display, Paris, France.

James, W. (1890). “Attention”, The principles of psychology (Vol. 1), Chapter 11, Holt, New York, USA.

Kahneman, D. (1973). “Attention and Effort”. Englewood Cliffs, NJ, Prentice-Hall.

Kleirock, L. (1996). “Nomadicity: Anytime, Anywhere in a Disconnected World”, Mobile Networks and Applications**,** Special Issue on Mobile Computing and System Services, vol. 1, n° 4, J.C. Baltzer AG, Science Publishers, p. 351-357, December.

McGee, M. R., Gray, P. D. and Brewster, S.A. (2000). “The Effective Combination of Haptic and Auditory Textural Information”, Proceedings of the Haptic Human-Computer Interaction 2000, First International Workshop, Glasgow, UK, p. 118-126, August.

Murphy, E., Kuber, R., Strain, P., McAllister, G. and Yu, W. (2007). “Developing Multi-modal Interfaces for Visually Impaired People to Access the Internet”, Proceedings of the 13th International Conference on Auditory Display, Montreal, Canada.

Nadin, M. (2003). “Anticipation - The End Is Where We Start From”, Lars Müller Publishers, Baden, Switzerland.

Oppermann, R. and Specht, M. (2001). “Contextualized Information Systems for an Information Society for All”, Proceedings of HCI International 2001, The 9th International Conference on Human-Computer Interaction, New Orleans, USA, p. 850-854, August.

Pashler H. (1995). “Attention and Visual Perception: Analyzing Divided Attention”, Visual Cognition, Chapter 2, Stephen Michael Kosslyn, Daniel N. Osherson (Eds.), p. 71-99, MIT Press.

Pashler, H., Johnston, J. C. and Ruthruff, E. (2001). “Attention and Performance”, Annual Review of Psychology, vol. 52, Palo Alto, CA, USA, p. 629-651, Annual Reviews.

Santaella, L. (2004). “O Método Anticartesiano de C. S. Peirce”, Editora UNESP, São Paulo, SP, Brasil.

Saussure, F. (1910). “Curso de Linguística Geral (Cours de Linguistique Générale)”, Editora Cultrix, 2006, São Paulo, SP, Brasil.

Searle, J. (1979). “Expressão e Significado (Expression and Meaning)”, Martins Fontes, 2002, São Paulo, SP, Brasil.

Treisman, A. and Gelade, G. (1980). “A Feature Integration Theory of Attention’, Cognitive Psychology, n° 12, 97-136.

Walker, B. N. and Nees, M. A. (2009). “Theory of Sonification”, In T. Hermann, A. Hunt, & J. Neuhoff (Eds.), Handbook of Sonification, New York: Academic Press, in press.

Want, R., Hopper, A., Falcão, V. and Gibbons, J. (1992). “The Active Badge Location System”, ACM Transactions on Information Systems, vol. 10, n° 1, January, p. 91-102.

Winograd, T. (2001). “Architectures for Context”, Human-Computer Interaction**,** vol. 16, n° 2, L. Erlbaum Associates Inc. Hillsdale, NJ, USA, p. 401-419, December.