

Estabelecendo uma Metodologia de Avaliação Subjetiva de Qualidade de Áudio

Marcelo Johann, Lucas Folle, William Gomes, Ricardo Cordoni

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

{johann.marcelo, follelucas, williamgomes1993, rocordoni}@gmail.com

***Abstract.** This paper reports the process of establishing a methodology for double blind AB or ABX tests for subjective audio quality evaluation. A series of different experiments were performed trying to identify audible differences between several analog/digital converters. We report both the failures as well as the final successes with statistical confidence levels, in order to highlight the difficulties as well as the potential of the technique. The main contribution of this work lies in the documentation of the process of setting up the methodology, identifying which parameters were necessary to make it work. One of the first conclusions, given the variety of results, is that tests in which individuals are not able to distinguish some feature do not prove that it is undistinguishable, but only that the test failed to identify that. We present a set of partial conclusions about the converters tested, specially regarding the limited quality of expensive "professional" products.*

***Resumo.** Este trabalho relata a experiência de estabelecer uma metodologia de testes de qualidade subjetiva de áudio, duplamente cegos tipo AB ou ABX, para avaliação de conversores analógico/digitais. Uma série de experiências foram feitas na tentativa de identificar diferenças audíveis entre vários equipamentos. Relatamos tanto os insucessos como os resultados finais positivos, isto é, com a capacidade de distinguí-los, para salientar a dificuldade e o potencial desse tipo de experimento. A principal contribuição desse artigo está na documentação da construção desse processo, identificando os parâmetros que foram necessários para que tivesse êxito. Conclui-se principalmente, pela variedade de resultados, que testes onde os indivíduos não conseguem perceber uma diferença não provam que ela não é perceptível, mas apenas que o teste falhou em identificá-la. São apresentadas as conclusões parciais que foram observadas com os conversores testados, evidenciando a qualidade relativamente limitada de equipamentos profissionais de alto custo.*

1. Introdução

Algumas décadas se passaram desde o surgimento do áudio digital e da proposta do primeiro formato popular chamado de Compact Disc, desenvolvido pela Phillips e Sony. E nesse meio tempo a computação passou de uma tecnologia empresarial e científica para se tornar pessoal, onnipresente, tanto para consumo como para geração de conteúdo. Percebe-se que atualmente uma parte considerável das atividades humanas são feitas com o auxílio, através, ou dentro do universo digital de computadores, com

instrumentos que contém processadores digitais, e a música não é exceção. Entretanto, apesar de ouvirmos música quase que exclusivamente no formato digital, na sua origem, geração, a valorização dos instrumentos e processos naturais, acústicos, parece nunca ter sido tão grande. Em outros termos, ouvimos música digital, em um reproduzidor MP3, ou em um DVD, ou pela Internet, mas a música que ouvimos é produzida com intenso uso de instrumentos acústicos, violões, guitarras, vozes, pianos elétricos, instrumentos originais que são preferidos em relação às melhores simulações digitais (*plug-ins* ou instrumentos virtuais). Todas as tecnologias digitais de que dispomos estão sim presentes, como caracteres adicionais, ou para conferir estéticas mais plásticas, modernas, diferenciadas, mas os músicos em geral reconhecem o valor insubstituível dos instrumentos acústicos, e isto não poderia, nem deve, ser diferente.

Entretanto, a popularização das tecnologias digitais nos oferece a oportunidade de difundir e estimular a produção independente de conteúdo, e especialmente a produção musical, democratizando-a (Valladares, 2011). Com dispositivos poderosos, de baixo custo e omnipresentes, o acesso ao estudo, à cultura e a produção musical não precisa mais depender de grandes instituições, estúdios e instalações, não precisa ficar sob os filtros e o controle de organizações com poder econômico. O conceito de música ubíqua surge então para descrever e estudar esse fenômeno: a atividade musical sendo realizada em qualquer lugar, por qualquer pessoa, com o auxílio de dispositivos digitais facilmente acessíveis, como aparelhos celulares, tablets, etc... (Pimenta et al., 2009) Apesar desse enorme potencial, também são muitos os desafios. Grande parte deles diz respeito ao desenvolvimento das tecnologias de *hardware* e *software*. Problemas com conexões adequadas de entrada e saída para dispositivos móveis (sua inexistência ou indisponibilidade) e problemas de latência são dois exemplos típicos. Um outro reconhecido problema (mesmo que informalmente) diz respeito à limitação de qualidade sonora obtida do som digital (ou digitalizado) quando amplificado e reproduzido por alto-falantes, em relação à riqueza e naturalidade do som acústico. O problema de produzir amplificadores e caixas de som de alta-fidelidade é antigo e bem conhecido (Olson, 2005), mas assume uma nova dimensão quando contracenado com o fator de disponibilidade: além de tudo, agora é necessário que o equipamento seja barato e extremamente pequeno, idealmente invisível, embutido nos celulares, TVs, etc.

Neste trabalho, consideramos uma parte importante desse problema, que é a qualidade subjetiva dos equipamentos de digitalização e reprodução digital, os conversores A/D e D/A. Embora a qualidade de amplificadores e caixas acústicas, como descrito acima, tenha um peso significativo, a qualidade subjetiva dos circuitos de conversão analógico digital parece ter um impacto muito grande na percepção musical. Por qualidade subjetiva entende-se a avaliação feita com sujeitos humanos, ouvintes, e não pelas medidas conhecidas de caracterização dos equipamentos musicais, como resposta de frequência, ruído, distorção harmônica total. A razão para o interesse em avaliação subjetiva, cuja motivação é melhor explicada na próxima seção, vem da observação de que equipamentos com parâmetros básicos muito bons, e semelhantes, parecem transmitir sensações musicais bem diferentes. Não se pode afirmar que não existem ou não são conhecidas as medidas que poderiam identificar essas diferenças. Mas isso não é fácil, e o objetivo final é sempre a audição, a satisfação da nossa percepção, que, em algum momento precisa ser considerada como parâmetro definitivo. É interessante notar que testes de qualidade subjetiva, com usuários, são e continuam sendo essenciais no desenvolvimento de sistemas e padrões de compressão de áudio. A

compressão com perda de informação é explicitamente a aceitação de um sistema com uma qualidade objetiva reduzida, ou a escolha dentre a perda de um ou outro tipo de informação, mas que subjetivamente tenha um resultado aceitável, ou perceptualmente melhor. Por isto rigorosos métodos foram desenvolvidos e são intensamente empregados nessa área para avaliar as novas tecnologias de compressão (normas ITU 1994-1997, 2001-2003).

Em testes de qualidade subjetiva, um conjunto de indivíduos deve experimentar o equipamento ou formato e responder a uma ou mais questões que visam verificar sua capacidade de distinção, sua preferência, classificar a qualidade percebida em uma escala ou em relação a determinadas referências, ou documentar suas impressões a respeito do sistema. Para eliminar polarizações, tendências, os testes devem ser feitos em ambientes controlados, nos quais nem o ouvinte e nem o cientista que controla possuem o conhecimento de qual o equipamento ou material é executado. Testes assim são chamados de duplamente cegos, e visam eliminar qualquer possibilidade de pré-disposição ou quaisquer dicas que possam influenciar nas preferências ou características percebidas. Os testes também devem ter validade estatística, e precisam ser repetidos um determinado número mínimo de vezes para que os resultados possam ser distinguidos de eventos aleatórios com determinado grau de certeza. Para satisfazer um grau de significância de 95%, são necessários 9 acertos em 10 tentativas, ou 15 acertos de 20 tentativas, conforme (Wikipedia, 2013 e Carlstrom, 2011). Muito conhecimento existe sobre esse tipo de experimento, com especialistas, centros e laboratórios dedicados, caros equipamentos de referência, e literatura (Bech, 2006).

Apesar dessas metodologias já bem estabelecidas, existem diversas questões recorrentes e oportunidades para verificar e melhorar a qualidade de equipamentos de áudio, como apresentamos a seguir, especialmente se a questão de custo também for considerada, mas não exclusivamente. Particularmente no Brasil, a engenharia de áudio é muito pouco desenvolvida, e metodologias de teste assim são pouco empregadas. Por outro lado, entendemos que o impacto artístico e musical é enorme. Ou seja, como pressuposto neste trabalho consideramos que as diferenças de qualidade resultantes do uso de diferentes conversores, ou de diferentes componentes em um equipamento, devem ser audíveis no produto final, mesmo no contexto de consumo em massa, ou de música ubíqua, (re)produzida em notebooks através de fones de ouvido. Assim, se por um lado pretende-se verificar nuances de qualidade em equipamentos do estado da arte, para os quais a infra-estrutura dos melhores estúdios e laboratórios de pesquisa pode ser limitante em termos de qualidade subjetiva, também consideramos que essa qualidade subjetiva tem um impacto direto na música e no usuário final, e deve ser percebida com equipamentos simples, e reproduzida em diversos lugares.

O restante do artigo está assim organizado. Na próxima sessão é apresentada a motivação do trabalho, tanto no contexto de interesses e aplicações locais, como em relação a questões recorrentes que surgem pelas diferentes visões das comunidades de engenharia e de artes (ou usuários), e na sessão seguinte são definidos então objetivos gerais e específicos. A seção 4 relata a série de experiências que foram feitas na tentativa de identificar diferenças audíveis entre vários equipamentos, apresentando principalmente os insucessos e os procedimentos que foram sendo tomados, permitindo gradual evolução. A metodologia que efetivamente nos permitiu distinguir equipamentos diferentes é então apresentada na sessão 5. Finalmente, as seções 6 e 7 apresentam respectivamente as principais conclusões e perspectivas do trabalho.

2. Motivação

A motivação de pesquisar sobre qualidade subjetiva de áudio, especialmente de conversores analógico/digitais, vem de aplicações práticas locais e experiências pessoais prévias. Porto Alegre tem um cenário cultural bastante rico tanto em termos de música popular quanto erudita. A Orquestra Sinfônica de Porto Alegre (OSPA) foi fundada em 1950, e tem portanto, mais de seis décadas de destacada atuação, com apresentações que incluem frequentemente convidados internacionais na posição de regentes e solistas. Além das OSPA, existem na grande Porto Alegre alguns outros grupos musicais muito importantes, como a Orquestra da ULBRA, a Orchestra do Teatro São Pedro, o Coral da PUC, o Coral da UFRGS, entre vários outros grupos musicais menores. Neste contexto, seguindo desde muito cedo os princípios de áudio de alta qualidade que foram desenvolvidos nas décadas de 1930 a 1950 principalmente na Inglaterra, o odontólogo e aficcionado por eletrônica e por música Marcello Sfoggia veio a se estabelecer como um excepcional perito em reprodução de áudio, construção de equipamentos de alta-fidelidade, e finalmente, especialista em gravações, mais especificamente em captação sonora acústica (Kafruni, 1996). A figura do Dr. Sfoggia, como ele mesmo conta, vem suprir uma enorme deficiência de profissionais de áudio qualificados, e principalmente dedicados. Assim como o inglês Gilbert Briggs (Briggs, 2012), considerado como pai da alta-fidelidade, e que não era engenheiro por formação, a paixão e dedicação do Dr. Sfoggia lhe permitiu dominar com maestria a construção de equipamentos de áudio, principalmente equipamentos analógicos valvulados (Sfoggia, 1990), mas também transistorizados e integrados (com amplificadores operacionais). Pela seleção de cada componente, seguindo o princípio do minimalismo, entre outros, e com décadas de montagens e comparações atentas, ele construiu um conjunto incomparável de pré-amplificadores, mesas de mixagens (Johann, 2013), entre outros, equipamentos todos dedicados (customizados), acompanhados de suas respectivas fontes de alimentação especializadas, com os quais fez rotineiramente o registro cultural da maioria absoluta dos grupos musicais (notadamente eruditos) da cidade, ao longo das últimas décadas.

Com o advento das gravações digitais, e a redução de custo dos novos meios de gravação, para melhorar a praticidade, as gravações regulares (série de todos os concertos da temporada de uma orquestra) passaram a ser feitas em fitas de vídeo e posteriormente em fitas digitais com os gravadores Alesis ADAT. Mais ou menos em 2008, com os gravadores ADAT já bastante obsoletos em termos de tecnologia mecânica e formato, e também em termos de desempenho dos circuitos conversores, era necessário passar para a já madura forma de gravação em computadores convencionais. Para tal, tivemos a oportunidade de estabelecer uma sólida colaboração no processo de selecionar um conversor analógico digital adequado. Os requisitos naturais eram que tivesse uma boa sonoridade, o mais parecido com o antigo ADAT, e idealmente mais próximo de um gravador analógico. Mas logo de início surgiram outras diretrizes que contradizem o senso comum, e mesmo alguns aspectos de consenso em engenharia: i) que marcas famosas profissionais não apresentavam um som satisfatório; ii) que aparelhos de baixo custo frequentemente apresentam uma sonoridade melhor, às vezes por coincidência, às vezes por terem menos componentes; iii) que existe um grande grupo de "modificadores" (modders), variando desde hobbistas (DIT - Do-It-Yourself) até engenheiros experientes que modificam fazem equipamentos dedicados para as aplicações mais importantes (Black Lion Audio, 2013; Newton, 2009).

Dentre um grande espectro de conversores, além das diversas considerações de recursos, conectividade, estabilidade, pode-se distinguir grosseiramente três grupos: 1) conversores populares, os mais baratos, na faixa de US\$150 a US\$500; 2) conversores profissionais de marcas renomadas, como Apogee, RME, Metric Halo, na faixa de US\$1000 a US\$2000; 3) conversores de ponta, usados apenas nas aplicações mais críticas, e de custo extremamente elevado, praticamente inacessíveis, com preços que podem exceder US\$8000 no mercado exterior. As diretrizes citadas anteriormente, portanto, traduzem-se em que não seria satisfatório usar conversores da segunda categoria, quando os produtos mais simples, com algumas modificações, podem resultar em melhor resultado sonoro. De forma resumida, durante essa experiência pessoal de selecionar um conversor (não era então um projeto acadêmico), foram seguidas quatro linhas de análise. a) Em primeiro lugar, análise de parâmetros técnicos, com uma lista de equipamentos, características (como frequência, número de bits, parâmetros publicados) e circuitos utilizados internamente; b) pesquisa por opiniões em fóruns e páginas que descreviam modificações, especialmente uma modificação proposta e relatada na interface Behringer ADA8000, que empregava os mesmos chips conversores do ADAT (Wavefront AL1201, AL1101); c) Testes e medidas com equipamentos disponíveis, que foram: uma interface Metric Halo ULN-2 atual (pertencente à categoria 2, de renome, comparável a RME e Apogee, com preço de lista de cerca de US\$1600), e as interfaces de baixo custo Behringer (US\$200), com e sem modificações; d) Finalmente, testes subjetivos utilizando somente a audição para comparação sonora.

Ao final foi dada muito mais atenção aos testes subjetivos, por uma simples razão: as medidas, o preço ou "renome" do equipamento de fato não explicavam as diferenças na impressão sonora que ele causava. Por exemplo, a resposta de frequência da interface Behringer, apesar de um pouco diferente da Metric Halo, é excepcionalmente boa, comparando-se com o que outros equipamentos de gravação típicos oferecem (microfones, mesas de som, etc). Mais importante ainda, após serem feitas uma série de modificações na interface Behringer (Johann 2010), e se ter a impressão de que o som havia melhorado significativamente, medidas como a resposta de frequência, com as características que haviam sido observadas, ficaram praticamente idênticas. A figura 1 mostra esta resposta com laços D/A/D nas interfaces Behringer original, modificada e na Metric Halo, nesta ordem de baixo para cima, com escala de 100dB (~16bits) e os detalhes presentes em 3dB. As curvas foram intencionalmente deslocadas com o controle de volume para poderem ser comparadas.

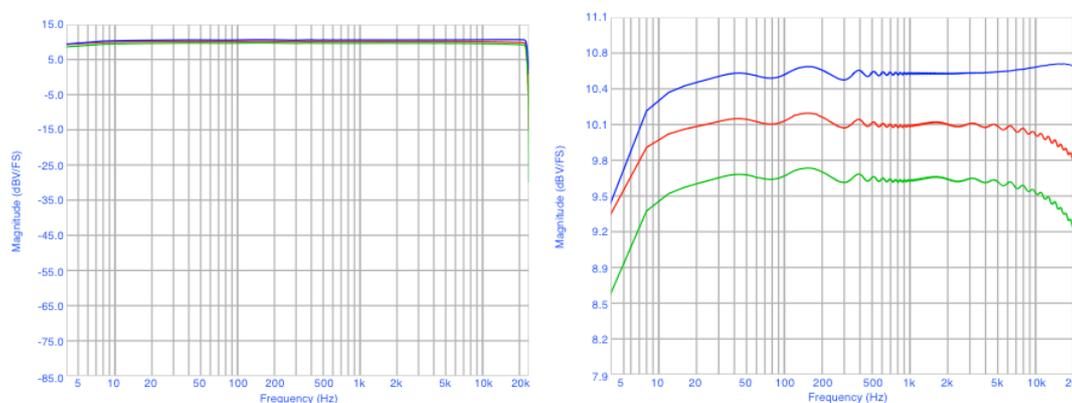


Figura 1. Resposta de frequência dos conversores Behringer Ada8000 original, com modificações e Metric Halo ULN2, em escala de 100db e detalhe de 3db

Ao final do processo, chegou-se à conclusão de que o resultado sonoro obtido com um conversor de baixo custo modificado, especificamente Behringer ADA8000, com o circuito pré-amplificador de microfone removido e amplificadores operacionais substituídos por outros de alto desempenho (Linear Technology LT1359 e Burr Brown OPA4134), era muito melhor, de forma geral, do que o som produzido por unidades caras de renome. Na tentativa de explicar o contexto técnico, econômico e sonoro que justifica esse fenômeno, uma apresentação/mini-curso sobre qualidade de áudio foi elaborado e apresentado em diferentes oportunidades (Johann 2011), e de fato, algumas razões puderam ser identificadas, e estão documentadas nessa apresentação.

Desde então, esse tipo de conversor modificado vem sendo usado para efetuar praticamente todas as gravações de música erudita em Porto Alegre. A escolha do conversor ADA8000 para modificação se dá principalmente porque sua arquitetura é simples, sua placa fácil de ser compreendida, mas pode-se obter resultados semelhantes com outros produtos, de diversas categorias. Entretanto, esse processo não acadêmico não possuía qualquer garantia ou validade científica. Foi feito exclusivamente com longas observações, audição, com a concentração e desenvolvimento da sensibilidade, que 'aparentemente' transmitia a sensação de que o resultado estava bastante diferente, e bastante melhor. A relatividade de tal conhecimento é evidente, e na nossa própria experiência variava de momentos onde as diferenças pareciam enormes, a momentos onde se ficava confuso sem saber qual equipamento ou circuito estava sendo usado.

A principal motivação para este trabalho então é muito clara. Deseja-se comparar conversores ou suas versões e modificações, em um ambiente controlado, medindo as opiniões com significância estatística e chegando a conclusões inequívocas sobre a nossa percepção de sua qualidade. Idealmente deseja-se identificar qual o conversor mais fiel, ou seja, que grava/reproduz o som mais próximo da fonte original, mas como será ressaltado, isso nem sempre é possível, e precisaremos contar com referências e conclusões intermediárias ou indiretas. Ao desenvolver a metodologia necessária para tais experimentos, poderemos então não somente validar as conclusões que já pensamos ter, como comparar outros equipamentos e opções de forma mais eficiente.

Além das aplicações locais, existem motivações científicas e sociais mais abrangentes. Documentar tais conclusões pode ser um primeiro passo para posteriormente entender e explicar porque um equipamento tem características que percebemos como melhores. Isso provavelmente implicará em selecionar medidas, desenvolver novas medidas, ou mesmo novas metodologias de teste em si. Não se pode ignorar que os engenheiros que projetaram os circuitos e produtos acima citados detêm enorme conhecimento, seguiram boas práticas, e empreenderam seu melhor esforço para produzir equipamentos com os melhores parâmetros, acreditando que tais medidas correspondem fielmente à qualidade sonora de seus produtos. Se isso não ocorre por qualquer razão, e de fato, há muitas opiniões controversas registradas, é necessário conduzir mais atividades de pesquisa que expliquem o fenômeno. Assim, indiretamente pode-se ter em mente outros objetivos maiores no futuro, como considerar as questões recorrentes de qualidade subjetiva evocadas em múltiplos debates.

3. Objetivos Específicos e Gerais

De acordo com as motivações detalhadamente descritas acima, o objetivo específico principal deste trabalho é comparar a qualidade de conversores AD e DA com testes subjetivos de audição, em ambiente controlado e com significância estatística.

Como objetivos mais gerais, a serem considerados no futuro, pretende-se também responder a outras questões recorrentes, na tentativa de explicar porque engenheiros e músicos possuem opiniões tão contraditórias, gerando inúmeros debates e conflitos. Para tal, procura-se desde já elaborar hipóteses de limitações nas metodologias de avaliação empregadas pela engenharia e ciência, as quais geram resultados questionados por usuários e músicos. Entre as principais hipóteses listamos inicialmente:

- 1- Que os testes controlados são feitos com trechos de áudio muito curtos, enquanto a percepção sonora musical será diferente quando o ouvinte apreciar o som por longos e relaxados períodos de tempo;
- 2- Que os equipamentos de referência usados em testes científicos não correspondem aos equipamentos culturalmente preferidos em ambientes musicais, o que influenciaria negativamente em testes de sensibilidade, isto é, faria com que os ouvintes não pudessem demonstrar adequadamente suas capacidades e preferências em ambientes controlados;
- 3- Que generalizações muito grandes são feitas. Dois exemplos são os seguintes. i) Se um experimento não funcionou, interpreta-se que foi demonstrado que outros não funcionarão. ii) Se ouvintes não conseguem distinguir a presença de um único componente X, ele não causará nenhum efeito sonoro, mesmo quando empregado diversas vezes em vários lugares ou equipamentos na sequência de produção musical.

Evidentemente existem outras possibilidades, outras variáveis psico-sociais que podem influenciar tais testes. Além disso, é importante notar que este trabalho é apenas o princípio de nosso projeto de pesquisa neste sentido, e não pretende responder a todas estas questões, que são aqui mencionadas por contextualização.

4. Experimentos Iniciais e Observações

Para reproduzir, entender e experimentar exploratoriamente a capacidade de distinção de diferentes conversores, foram feitas três sessões de testes em um auditório nos dias 25/05/2013, 08/06/2013 e 06/07/2013. No primeiro dia, procuramos ouvir trechos de gravações com diferentes circuitos de conversão DA. Usando uma chave manual para mudar de um conversor para outro, não pudemos obter nenhum indício de capacidade de distinção, com resultados praticamente aleatórios e em poucos testes. Não somente a capacidade de acertar na identificação de conversores não foi demonstrada, como nossa impressão geral também foi que nenhum era diferente ou superior, mesmo quando ouvindo as gravações informalmente. De fato, essa dificuldade, que foi maior do que a expectativa, persistiu parcialmente nas próximas duas sessões, como segue.

Em todos os casos, as audições e testes foram feitos com equipamentos de razoável qualidade tanto técnica como musical, sendo o som amplificado e ouvido no ambiente. A sala possui apenas um tratamento acústico básico, por se tratar de um auditório, mas não é caracterizada e nem possui tratamento especial com qualidade de

estúdio ou científica. Foram usados um par de monitores Alesis Monitor One MkII e um par de monitores Event 20/20, todas passivas, ligadas a um amplificador Nashville Power-70, com entrada adaptada para conexão DC.

Nas duas sessões seguintes, passamos a usar teclados sintetizadores digitais como fontes sonoras, para testar então o laço AD/DA, isto é, partindo de um sinal analógico (resultante dos teclados), que é digitalizado e depois volta ao domínio analógico para audição. Isso foi feito por três razões. Primeiro, pois correspondia à forma de teste usada originalmente, conforme relatado na seção 2 e em (Johann, 2010). Segundo, porque dessa forma o teste deve ter um potencial maior, pela comparação do laço, em vez de apenas DA, e porque compara-se com o original, e não um DAC com outro DAC. Em terceiro lugar, o teste com um teclado é mais prático pois temos o controle de tocá-lo de forma instantânea, e inclusive a possibilidade de o músico (que o toca) ser influenciado pelo som do instrumento. Essa última característica requer muito cuidado, pois pode ser uma forte influência tendenciosa para outro ouvinte, mas teria o potencial de amplificar a percepção quando o ouvinte é apenas o músico que toca.

Os experimentos com teclados nessas duas sessões também foram muito difíceis. Em grande parte, as respostas para identificar a diferença entre um equipamento e outro não superavam entre 4 a 6 acertos de 10, em diversas configurações. Esses resultados não foram anotados, não pelo insucesso, mas por estarmos em uma sessão explicitamente exploratória, para verificar o que era possível e quais os problemas. Entretanto, observamos uma sequência de avanços significativa, bem como aspectos que deveriam ser isolados.

No dia 08/06, em um teste final, um dos participantes demonstrou a capacidade de distinguir o som através do DAC da RME (conectado digitalmente ao teclado) da sequência composta pelo DAC do teclado (Yamaha Motif) sendo depois digitalizado e voltando ao domínio analógico pelos conversores de baixo custo da Behringer ADA8000 original, com 10 acertos de 10 tentativas, após ter selecionado cuidadosamente um dos timbres (de *combo-organ*) e fazer o teste bem rápido. No dia 06/07, também após ter grande dificuldade, foi possível apenas distinguir com acertos entre 8/10 a 10/10 a diferença entre o som analógico do mesmo teclado e ele passando novamente pelo laço A/D/A da ADA8000. Mas não foi possível distinguir os laços digitais através da ADA8000 e dos produtos mais caros da RME e Metric Halo, que consistem em um ponto de partida básico, conforme a motivação.

Em relação à dificuldade, identificamos vários aspectos e possibilidades. A acústica da sala não era adequada. A execução dos sons, feita manualmente, tinha grande influência, e mesmo a percepção fica confusa, pois se presta atenção a diferentes características. Nem sempre o que chama mais atenção é o melhor, nem sempre o que parece mais brilhante, ou mais presente é o melhor, já que distorções amplificam ou geram frequências, conferindo características mais marcantes ao sinal. Observamos também falta de foco e cansaço. Comentários, ou a necessidade de falar para mudar a chave manualmente, faziam com que a atenção não estivesse exclusivamente dedicada para perceber as diferenças. Diversos problemas foram eliminados, como a sensibilidade física do teclado, a existência de timbres com múltiplas amostras por intensidade, as funções de panorama aleatório e efeitos como chorus, entre outros.

5. Metodologia Efetiva e Resultados

Considerando a grande dificuldade encontrada nos testes com o som sendo amplificado, mesmo com a possibilidade de comparar os conversores com o som "original", direto da fonte para o amplificador, passamos a uma segunda fase, que foi a comparação de áudio digitalizado com diferentes conversores, e avaliado com fones de ouvido. Esta foi a segunda opção, pois tem uma limitação intrínseca, a de que o resultado sonoro será limitado, ou condicionado, ao DAC que for usado para audição. Ou seja, não se pode comparar simultaneamente o laço AD/DA de dois equipamentos, nem compará-lo com o original, sem envolver uma automação de conexões analógicas ainda indisponível. Assim, inicialmente não se esperava que esse tipo de comparação fosse mais fácil.

Mas ao contrário, mesmo com essa limitação, para teste dos conversores AD apenas, essa metodologia se mostrou muito mais efetiva. Ela elimina a maior parte dos problemas experimentados nas sessões de audição, e permite testes muito mais rápidos, práticos, feitos em diversos equipamentos. Foram usados dois programas já disponíveis para testes ABX, ABXer para mac e Foobar2000 para Windows, e os arquivos de (Johann 2010). Mesmo assim, observamos inicialmente uma razoável dificuldade na identificação, com taxas de acerto entre 40% e 60% nas primeiras séries de 10 tentativas, o que é indistinguível de um evento aleatório.

Mas logo em seguida, identificamos um dos elementos mais importantes para a metodologia. O teste deve ser feito com trechos extremamente curtos, preferencialmente em um único ponto (uma nota ou um acorde), previamente selecionado. À partir da identificação desse requisito e da seleção de um ou outro ponto específicos nos arquivos de áudio, as diferenças passaram a ser identificadas com taxas de acerto de até 9/10 ou 10/10, por diferentes indivíduos e com equipamentos de audição (DAC e fone) de diversos padrões de qualidade. A tabela 1 abaixo mostra todos os resultados que foram registrados de nossos testes com fontes de ouvido para distinguir dois conversores AD: o da Metric Halo ULN2 e um dedicado, projetado localmente, usando o chip PCM4222 da Texas e um buffer de entrada com operacionais LT1359 e acoplamento DC. A escolha desse conversor se deu principalmente porque se buscou aumentar passo a passo o grau de dificuldade, e em experiências prévias havia informalmente a sensação de que esse conversor possuiria a sonoridade mais próxima do som original, o que é plenamente justificado por ele ser um circuito recente e o melhor conversor integrado produzido pela Texas Instruments.

Os resultados na tabela incluem tanto as séries de teste onde trechos de tamanho variado (ou o arquivo completo) foram ouvidos, como aquelas restritas a um ponto de 1 a 2 segundos que começa em 00:50 no arquivo, apresentados na segunda metade da tabela. Não foram registrados resultados considerados como de treinamento. Ou seja, os indivíduos tiveram acesso aos arquivos e podiam tanto anteriormente quando durante o teste, conforme os programas de teste permitem, ouvir quantas vezes quisessem cada amostra, antes de se decidirem a efetuar a série de opiniões. É interessante observar também que os programas de teste ABX também sorteiam inicialmente as referências A e B, o que seria apropriado para alguns testes, mas não necessariamente neste caso, onde o objetivo é demonstrar a capacidade de distinção ou identificação do equipamento, em primeiro lugar. Assim, havia uma dificuldade a mais: era necessário primeiro identificar qual conversor era o A e o B, para depois seguir a série de acertos no palpite de qual é o X, sorteado a cada iteração.

Tabela 1. Acertos na identificação dos conversores AD da Metric Halo ULN2 e projeto local com circuito PCM4222 da Texas Instruments

Length	Listeners / Correct				DA Converter					HeadPhone				Sum	
	I1	I2	I3	I4	RME	MH UlN2	iMac	Fast Tr	Sony	HD650	K270	V150	SHP2k		Phil/Koss
any			4				X						X		4/10
any			4				X						X		4/10
any			8				X						X		8/10
any			7				X			X					7/10
any			7				X			X					7/10
any			8				X			X					8/10
any	6	8			X		X			2X					14/20
any	7	6			X		X			2X					13/20
any	5	6	5			X	2X			X	X		X		16/30
any	6	6	5			X	2X			X	X		X		17/30
any	7	8	4			X	2X			X	X		X		19/30
<2s	10			10	8	X	X	X		X	X			X	28/30
<2s	9			8	8	X	X		X	X	X			X	25/30
<2s	8			9	9	X	X	X		X	X			X	26/30
<2s	9			8	7	X	X		X	X	X			X	24/30
<2s	9			6		X	X			X	X				15/20
<2s	10			9			2X			X	X				19/20
<2s	9			8			2X			X	X				17/20
<2s	10			10			2X			X		X			20/20
<2s	10			8			2X			X		X			18/20
<2s				10			X			X					10/10
<2s				10			X			X					10/10
<2s				10			X			X					10/10
<2s				10			X			X					10/10
<2s				10			X			X					10/10
<2s				9			X			X					9/10
<2s				6			X			X					6/10
<2s				9			X			X					9/10
Hits	115	72	164	32											383
Totals	140	110	200	40											490
Random probab.	2.9 e-15	0.07%	<9.9 e-16	9.1 e-5											<9.9 e-16

O número de testes e os equipamentos não foram uniformes nesta etapa do trabalho, para estabelecer a metodologia. Mesmo assim, os resultados gerais e de cada indivíduo possuem grande significância estatística. A última linha da tabela indica a probabilidade de um evento aleatório produzir esse resultado, e é em todos os casos muito inferior a 1%. Pode-se observar também que após identificarmos e concentrarmos a audição a um trecho de 1 ou 2 segundos, a grande maioria das séries individuais de 10 tentativas satisfaz o número mínimo de acertos (em negrito). Pode-se afirmar então com grande grau de certeza que os ouvintes percebem a diferença de qualidade destes dois conversores. Mesmo quando os testes empregaram conversores DA e fones de mais baixa qualidade, as diferenças puderam ser percebidas, mas mais experimentos devem ser feitos para aferir a dependência ou independência desses sistemas.

6. Conclusões Parciais

As principais conclusões obtidas neste trabalho foram as seguintes:

- Um experimento com resultados que não se distinguem do aleatório, como nas primeiras sessões de audição, não indicam que a percepção não existe, apenas que o experimento falhou em identificá-la;
- Os testes precisam ser automatizados, rápidos e permitir total concentração e foco;
- O uso de trechos bem curtos, uma única nota ou acorde, entre 1 a 2 segundos, previamente selecionados, permite mais facilmente identificar diferenças;
- Considerando todos os testes feitos, e para cada ouvinte, demonstra-se a capacidade de identificar as diferenças entre os conversores Metric Halo ULN2 e o projeto local com o circuito PCM4222;
- As diferenças podem ser percebidas até mesmo em equipamentos de consumo e baixo custo.
- A identificação dos conversores esteve diretamente associada à superioridade do PCM4222, que tem mais detalhes, e deve, ser mais próximo do som original;

7. Direção para Trabalhos Futuros

Já que a metodologia aqui desenvolvida é restrita ao universo digital, os próximos passos incluem o projeto da automação de chaves, que permitirá comparações no domínio analógico com a mesma estratégia. Os testes devem ser repetidos com outros indivíduos, equipamentos, hipóteses e variáveis de controle, para gerar resultados mais completos e expressivos. Espera-se futuramente usar a metodologia também para julgar qualidade de componentes e outras decisões de projeto de vários tipos de circuitos, para projetar equipamentos melhores, e também avaliar outros fatores que influenciam a sua execução, como experiência, treinamento, fadiga, ou outras variáveis que possam ser encontradas, de forma a ter testes mais detalhados e eficientes.

Agradecimentos

Este projeto recebeu suporte do CNPq através do Edital MCT/CNPq 14/2010 - Universal, processo 484423/2010-0 e dos programas de bolsas institucionais UFRGS/CNPq para iniciação científica e iniciação tecnológica. Agradecemos a Rodolfo Pirotti, Andrei Yefinczuk e Lucas Zawacki pela participação em sessões de testes como avaliadores (ouvintes) e com palpites críticos. Agradecemos também ao Dr. André Luis Korzenowski (Unisinos) pelo auxílio na análise estatística dos resultados.

Referências

- Bech, S. e Zacharov N. (2006) "Perceptual Audio Evaluation – Theory Method and Application", John Wiley & Sons Inc., 1ª edição.
- Briggs, David. (2012), "A Pair of Wharfedales - The Story of Gilbert Briggs and his Loudspeakers", IM Publications, 359p.
- Carlstrom, David. (2011) ABX Web Page - ABX Double Blind Comparator. Disponível em: <http://home.provide.net/~djcarlst/abx.htm>. Acessado em 15/09/2013.

- International Telecommunication Union (1994-1997) “Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems”, http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-1-199710-I!!PDF-E.pdf, Setembro.
- International Telecommunication Union (2001-2003) “Methods for the Subjective Assessment of Intermediate Quality Level of Coding Systems”, http://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1534-1-200301-I!!PDF-E.pdf.
- Johann, Marcelo. (2011) “Who Believes in Audio Quality” Disponível em: <http://www.inf.ufrgs.br/~johann/semac2011johannUFPel.pdf>, acessado em Set/2013.
- Johann, Marcelo. (2011) “Quem se Importa com Qualidade do Áudio, Afinal?”. 13 Simpósio Brasileiro de Computação Musical, Vitória, ES, 31-Ago a 03-Set de 2011, Tutorial. Resumo disponível em: "<http://compmus.ime.usp.br/sbcm/2011/pt/tutorial.html>".
- Johann, Marcelo. (2010) “Modifying the ADA8000”. Disponível em: "<http://www.inf.ufrgs.br/~johann/ada8000/>", Acessado em Setembro/2013.
- Johann, Marcelo. (2013), “Mesa de Gravação do Marcelo Sfoggia”, <http://www.youtube.com/watch?v=uQQ0-7N5fb0> , Acessado em Set/2013.
- Kafruni, Simone. (1996) “O Apreciado Som do Dentista Sfoggia - O odontólogo Marcelo Sfoggia é o mais requisitado técnico para as gravações de música clássica”, Correio do Povo, Domingo, 15 de dezembro de 1996.
- KVR Audio, <http://www.kvraudio.com/forum/>, Setembro.
- Newton J. (2009), Modified Nakamichi DMP 100, como visto em <http://www.youtube.com/watch?v=G6sTORRJv5s>, Setembro.
- Olson L. (2005), “A Tiny History of High Fidelity”, Disponível em: <http://www.nutshellhifi.com/library/tinyhistory1.html>, Acessado em Set/2013.
- Phillips e Sony (1980) “Red Book”, Philips e Sony, 1ª edição.
- Pimenta, M. S., Flores, L. V., Capasso, A., Tinajero, P. (2009) Ubiquitous Music : Concepts and Metaphors. *Proceedings*. XII Brazilian Symposium on Computer Music, p. 139-150.
- Sfoggia M. (1990), “Marcello Sfoggia talking about his audio equipment in 1990”, <http://www.youtube.com/watch?v=yOVVPFRx3bQ>, http://www.youtube.com/watch?v=27795_PAKP0, Setembro.
- Theobald, Ruth, “The 70 year history of Gilbert Briggs and his company - Wharfedale”. GearPlus, November 2007. Disponível em: "<http://www.gearplus.com.au/products/wharfedale/history/0-history-wharfedale.htm>", Acessado em 15/09/2013.
- Valladares, T. A. (2011). Independent Producers: A Guide to 21st Century Independent Music Promotion and Distribution (Master Thesis), Available at <http://hdl.handle.net/1794/11217>
- Wikipedia (2013), ABX test. Disponível em: http://en.wikipedia.org/wiki/ABX_test, Acessado em 15/09/2013.