

Como medir a qualidade de áudio?

Thilo Koch

Grupo de pesquisas em Computação Musical
IME - USP



18 de maio de 2016

Motivação

- Escutamos muitas fontes de som diariamente:
 - sons do ambiente na cidade, veículos nas vias (carro, buzina),
 - diversos estilos de música, arte,
 - fala em sala com ventiladores,
 - sons / barulhos e ruídos no ambiente de trabalho (máquinas),
 - condições de escuta (banda ao vivo, fones de ouvido, elevador, celular na rua / em casa),
 - condições mentais da escuta.
- Existe uma ideia de qualidade "boa" ou "ruim" especialmente na área de produtos.
- Recentemente academia e indústria pesquisam: como **quantificar a qualidade percebida de áudio**,
- Executar medições de qualidade (de maneira certa) é extremamente complexo como vamos ver neste seminário.

Introdução

Nesse seminário será apresentada uma introdução e visão geral da avaliação de áudio. Como planejar e executar experimentos, e analisar os resultados. Será um tipo passo a passo para demonstrar o processo de experimentação científico.

Introdução

Questão de pesquisa e hipótese

Preparação e execução do experimento

Quantificação da impressão

Variáveis independentes

Objetivo

Objetivo: Medir a qualidade percebida de estímulos auditivos.

- Problema multidimensional porque envolve o lado físico e o lado cognitivo.
- Medir apenas características físicas (intensidade, campo sonoro, etc.) não revela como o sistema auditivo interpreta e quantifica os estímulos.
- 2 alternativas: modelos preditivos de percepção (por ex. computacionais) ou teste de escuta (Perguntar ao ouvinte como foi.).
- Medição perceptual: perguntar sobre certas características auditivas do estímulo.
- Medição afetiva: perguntar sobre a impressão geral do estímulo.

Teste de escuta

O que pode ser esperado de testes de escuta quando à qualidade de estímulos auditivos?

- Identificar que dois estímulos são perceptualmente iguais.
- Identificar que um estímulo é igual, superior ou inferior a um outro estímulo em termos da qualidade percebida.
- Determinar o grau de diferença entre estímulos em termos da qualidade percebida.
- Com esses experimentos, é possível estabelecer escalas de vários níveis.
- Outros, por exemplo: Qual sistema de áudio é preferido.

O que não pode ser esperado de testes de escuta?

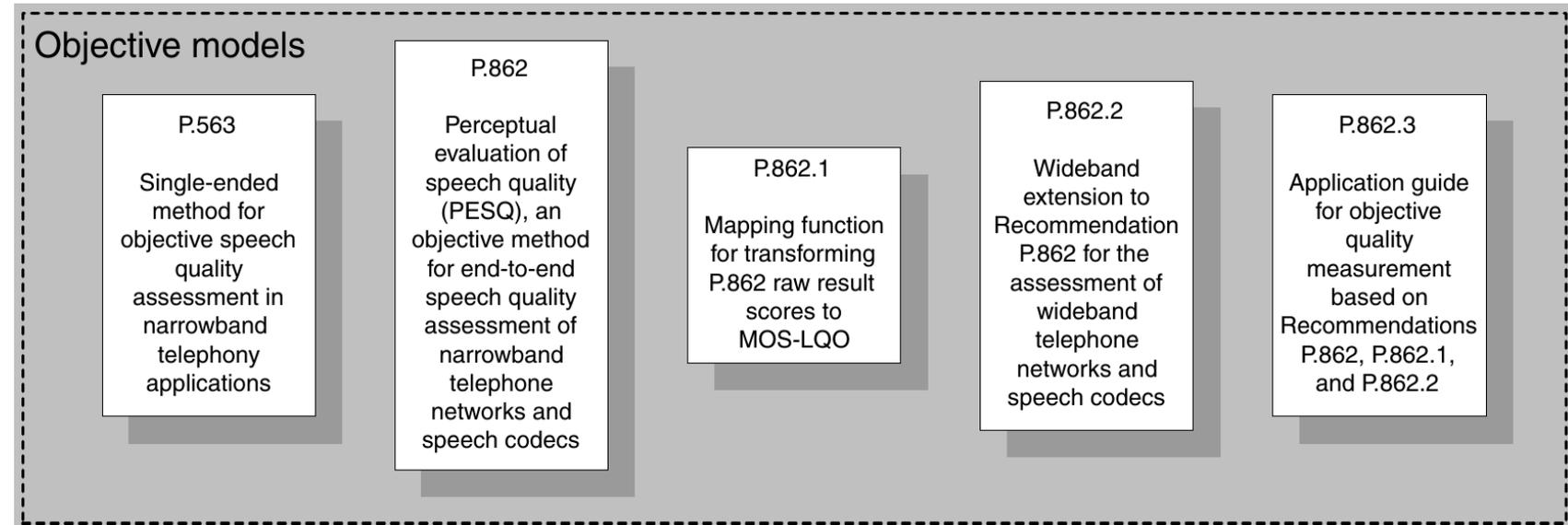
- Localizar parâmetros problemáticos em algoritmos de processamento.
- Como melhorar a percebida qualidade de áudio.
- etc.

Padrões, Normas e Recomendações

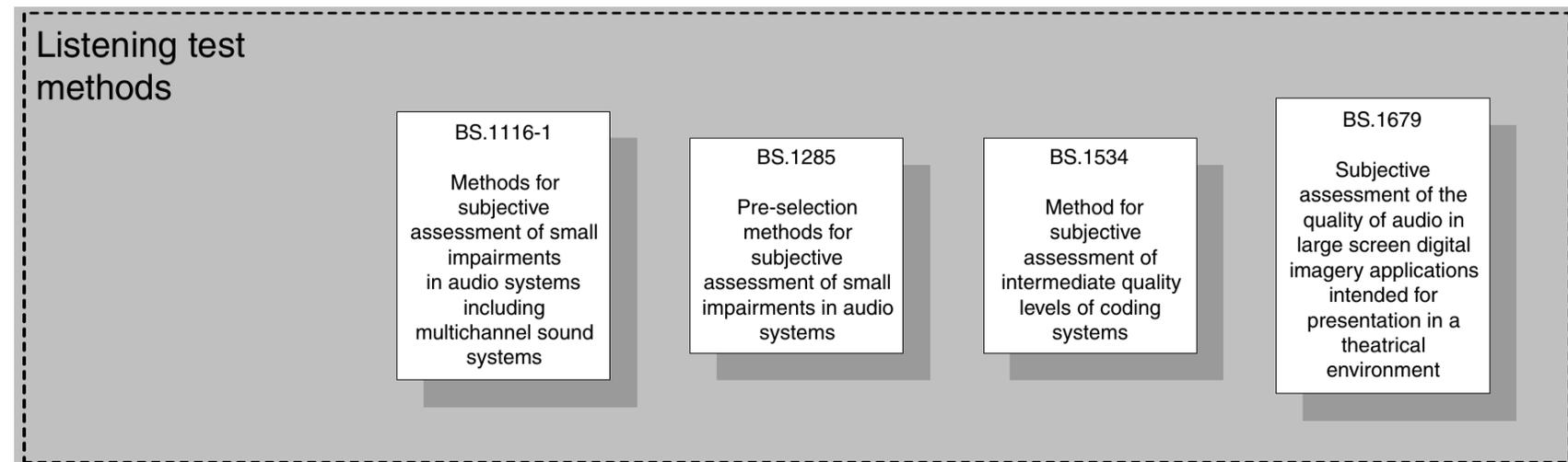
- Existem vários padrões para áreas diversas de vários órgãos:
 - Audio Engineering Society (AES),
 - American National Standards Institute (ANSI),
 - European Broadcasting Union (EBU),
 - International Telecommunication Union (ITU).
- São desenvolvidos quando se tem interesse ou necessidade de vários partidos.
- Vantagem: metodologias padronizadas são reconhecidas na academia e indústria
- Vantagem: Não tem que desenvolver uma metodologia e pode confiar na expertise que está embutida nesses padrões e normas.

Padrões, Normas e Recomendações

ITU-T: qualidade de aplicações de telefonia e de voz (*speech*) para escutantes não treinados;



ITU-R: áudio de banda larga, escutantes treinados; BS 1387: método para avaliação objetiva de qualidade de áudio percebida



Modelos preditivos de percepção

- Para prever uma característica particular (por exemplo *loudness*) ou desempenho geral (qualidade de escuta);
- normalmente contém um modelo do sistema auditivo acoplado a um modelo cognitivo.
- Vantagem óbvia: simplificar e acelerar testes de escuta.
- Else são válidas apenas em domínios particulares e definidos (telefonia, localização em uma dimensão, etc) - não podem ser usados fora desses domínios e condições.
- Definidos em padrões, recomendações, publicações científicas, etc
- Exemplos: avaliação perceptual de qualidade de voz (*speech*) para banda estreita em telefonia (ITU-T P.862), seleção de modelos de localização de fontes sonoras (Auditory Modeling Toolbox¹)

¹<http://amtoolbox.sourceforge.net>

Questão de pesquisa e hipótese

Contexto científico

Ciência produz novo conhecimento por experimentação testando hipóteses usando as normas e métodos estabelecidos. Esse conhecimento é verdadeiro por definição. (Conhecimento falso não é produzido.) Conhecimento científico é:

- teórico - porque é generalizado / abstração da realidade;
- verdadeiro - porque é produzido pelas normas e métodos estabelecidos;
- sistemático - baseado em conhecimento sistematizado de experimentos anteriores.

A questão de pesquisa é fundamental para a experimentação.

- Formulação abrangente do problema de interesse / hipótese
- Divisão / extração de questões que podem ser respondidas por experimentos

Como funciona o raciocínio que liga a hipótese ao experimento e ao conhecimento científico?

Princípio do Empirismo

- Generalizar do particular ao geral,
- de observações singulares em experimentos derivam-se leis gerais.
- **Inferência indutiva:**
 - Premissa: A é observado " muitas" vezes junto com B mas nunca sem B .
 - Conclusão: $A \rightarrow B$ (é válida / verdadeira)
- Problema: como se justifica a inferência indutiva logicamente? ???
por exemplo: Kant: Inferência indutiva é válida *a priori*.
- Crítica: Aceitar a inferência indutiva leva a inconsistências lógicas.
- Na prática experimental: procura por resultados afirmativos prejudica a produção de resultados desviados.
- Normalmente não procuramos observar B sem A (assimetria).

Princípio do Racionalismo

- **Método dedutivo de testar** (*the deductive method of testing*) - Karl Popper
- Não existem leis gerais mas teorias que foram comprovadas e conservadas até agora (*bewährt*).
 - *teoria1* : $A \rightarrow B$;
 - *condição1* : A é observado;
 - *prognose1* : B deve ser observado.
- Como se justifica o método? Pela lógica dedutiva.
- Na prática experimental: Tentar contraditar a negação da *teoria1*.
 - *teoria2* : $\overline{A \rightarrow B}$;
 - *condição2* : A é observado;
 - *prognose2* : B não deve ser observado.
- A partir disso: procura por observações que contraditam a *prognose2* para contraditar a *teoria2* e eventualmente "comprovar" / melhorar a *teoria1*

Procura por erros (trial and error) é um método mais rico porque erros são boas fontes de descobertas.

Exemplo

Questão de pesquisa: "Alto-falantes soam diferente em salas diferentes e em posições diferentes?"

Método empírico indutivo:

- Premissa: Foi observado muitas vezes que os mesmos alto-falantes em posições diferentes numa mesma sala, soam diferente.
- Conclusão: Para todos os alto-falantes é válido que o mesmo alto-falante soa diferente em cada posição diferente, numa mesma sala.

Método dedutivo:

- Teoria: Alto-falantes idênticos colocados em posições diferentes na mesma sala, soam **similar**;
- Condição: Dois alto-falantes idênticos em duas posições diferentes são comparados.
- Prognose: Os alto-falantes vão soar similar.
- Estratégia: Tentar "cumprir" a prognose, investigando e refinando a condição (testar alto-falantes diferentes). Obviamente a teoria é válida para alguns tipos de alto-falantes, e para outros não.

Processo de preparação do Experimento

- Definição da **hipótese**
- Estudo de **bibliografia** - aprender com outros: set-up experimental, técnicas, dados para a comparação, ...
- Existem **modelos preditivos perceptuais**? → Testes automatizados sem teste de escuta.
- Existe um **padrão** ou uma recomendação apropriada?
- Definir as **variáveis independentes e dependentes**:
 - as dependentes: resposta do sujeito (formato da resposta, escala, etc.)
 - independente: condições que o experimentador pode influenciar (características do sinal de teste, sistema de reprodução, situação de escuta, escolha de sujeitos)
- Definir **base estatística** do experimento: número de sujeitos, número de testes, número de repetições, variações de parâmetros relevantes.
- Definir **modelo estatístico**: Análise tem que determinar se a influência das variáveis independentes controladas às variáveis dependentes é mais que acidental ou realmente contribuem na variação das variáveis dependentes (p).

Fase experimental e pós experimental

- Executar um **experimento piloto**. Checar e caso necessário, refinar:
 - o set-up e a calibração;
 - se o experimento realmente testa a prognose (*testable statement*)
 - variáveis, parâmetros e faixas dos estímulos, modelo estatístico
- **Documentação** do set-up: pesquisa reprodutível (*reproducible research*)
- **Execução do experimento**
- **Análise** dos resultados - estatística com modelo já estabelecido.
- Conclusões e **discussão** e/ou reformulação da hipótese - voltar para o começo. :)

Quantificação da impressão

O cerne do Experimento:

- Sob certas condições (variáveis independentes),
- o sujeito (variável independente) é exposto a um estímulo (variável independente).
- O sujeito dá respostas (variáveis dependentes) às questões sobre sua percepção (*~ prognose*).

Nesta resposta está embutida a transformação - que o sujeito realiza - do estímulo percebido em um tipo e formato de saída. Essa **transformação do universo analógico das percepções do sujeito em valores discerníveis e discretos na resposta** é o fundamento da ideia que certas propriedades de áudio sejam quantificáveis.

Response attribute

- *Response attribute* é o "tipo" de resposta que o sujeito pode dar e é relacionado à propriedade do áudio percebido que deve ser quantificada.
- Por exemplo: *loudness*, *pitch*, posição espacial, ...
- Tem que ser explicada para o sujeito.

Estímulos complexos tem que ser analisados anteriormente. As propriedades tem que ser levantadas e formam o vocabulário consensual. Se não existir - no idioma correto - tem que ser desenvolvidos, normalmente em processos de grupo. Esses processos são muito custosos e divididos em várias fases cada uma de várias sessões (meses; recomendação: grupos permanentes). Esse processo de levantamento inclui:

1. identificar propriedades a partir de estímulos representativos;
2. definir e descrever as propriedades;
3. definir escalas para as propriedades;
4. testar a consistência e estabelecer uma *baseline* para os experimentos.

Esses vocabulários consensuais tem que ter certas propriedades, como precisão das palavras, pouca sobreposição, inambiguidade, capaz de gerar consenso.

Response format

- O formato da resposta inclui uma **escala** e um **método** que o sujeito usa para relatar a impressão.
- Um método padrão é **atribuir números a impressões** (em vez de usar textos individuais) porque é:
 - objetivo,
 - quantificável,
 - comunicável facilmente,
 - generalizável.
- *direct scaling* - o sujeito converte a sensação diretamente em uma magnitude e relata-la.
- *indirect scaling* - o sujeito tenta discriminar entre estímulos.
 - baseado na premissa: o número de vezes em que um estímulo é estimado diferente de um outro é diretamente ligado ao grau da diferença da sensação.
 - métodos comuns: comparar estímulos constantes, método de limites, método de ajuste

Níveis de escalas

Scale	Basic operation	Permissible transformations	Permissible statistics	Examples
Nominal	Equality versus inequality	Any one-to-one	Number of cases, mode	Telephone numbers, and so on
Ordinal	Greater than versus less than	Monotonically increasing	Median, percentiles, order statistics	Hardness of minerals, class rank
Interval	Equality of interval or differences	General linear $x' = b \times x + a$	Arithmetic mean, variance, Pearson correlation	Temperature (Celcius) (conventional test scores)
Ratio	Equality of ratios	Multiplicative (similarity) $x' = b \times x$	Geometric mean	Temperature (Kelvin)

Steven's level of measurement, basic defining operations, permissible transformations, example of permissible statistics, and examples. From [328]. J. Nunnally and I. Bernstein, *Psychometric Theory*, 3rd ed., 1994, McGraw Hill. Reproduced with permission of The McGraw Hill Companies.

Contexto e Vieses

Vieses tem que ser minimizados ou excluídos pela configuração do experimento.

Exemplos são:

- viés de contração: sujeitos subestimam grandes diferenças e sobrestimam pequenas diferenças;
- viés de contração sequencial: o estímulo anterior tem influência na percepção do estímulo seguinte;
- viés por causa de magnitudes (de diferença) pouco familiares: comparar o incômodo auditivo de um rato com o de um avião;
- viés na escalas: escala logarítmica com número diferentes de dígitos, problemas de não usar a faixa de valores inteira da escala, etc);
- outros: viés visual, pela espacialização, pela esperança, *cross modality*.

Variáveis independentes

- A configuração do experimento tem que garantir que a variação nas variáveis dependentes é causada pelas variáveis independentes e não por variáveis não conhecidas (e por isso, não são controláveis).
- Quanto menor o número de variáveis não conhecidas, melhor será a precisão dos resultados (por exemplo ANOVA).
- As variáveis independentes mais importantes são **o sinal, o sistema de reprodução, a situação de escuta e os sujeitos testados.**

O Sinal

- Pode ser qualquer sinal - sinais sintéticos (ruídos ou calculados) ou sinais "naturais";
- Tem que excitar as diferenças perceptuais em questão.
- É selecionado na base dos *response attributes*.
- Tem que ser apropriado quanto às propriedades temporais, espectrais e espaciais.

O Sujeito

O sujeito de teste é o instrumento de medição.

- As características dos sujeitos são importantes fatores para estimar a generalidade das conclusões do experimento.
- A seleção de sujeitos se baseia nessas características.
- Recomendações definem categorias de sujeitos (ITU-T, ITU-R): sujeito não treinado, sujeito experiente, *expert*, etc.
- Seleção randomizada tem que ser planejada e restringida (por exemplo critérios de representatividade).

O Objetivo da experimentação é ...

Atingir **Objetividade**: dados e conclusões similares podem ser obtidos por um outro experimentador usando os mesmos estímulos mas com outros sujeitos, sob outro paradigma de experimentação, em um outro lugar.

Ou pelo menos **Reprodutibilidade** que é como objetividade mas com o mesmo paradigma experimental.

Fim

Muito obrigado!

Perguntas? Comentários? Sugestões?

Contato: tiko@ime.usp.br