

# Medir o STI

Swen Müller

<sup>1</sup>Instituto Nacional de Metrologia, Normalização e Qualidade Industrial (INMETRO)  
Divisão de Acústica e Vibrações (DIAVI)  
Avenida Nossa Senhora das Graças, 50  
Xerém - Duque de Caxias - RJ, 25250-020 – Brazil  
smuller@inmetro.gov.br

**Abstract.:** *Worldwide, the STI is gaining increased importance for the evaluation of intelligibility in public places. Often, the STI already is part of the requirements in the solicitation for the construction of a new building. If a minimum stipulated STI value cannot be reached after finishing the construction, the building contractor has to rectify the problem, otherwise he is held liable and subject to contractual penalties. For this reason, the STI is already simulated in the planning phase and later verified by measurements on site. This article tries to present the fundamental concepts of the STI and the techniques and signal processing to obtain it. Some theoretical and practical examples are shown and some weak points of the STI revealed.*

**Resumo.:** *No mundo inteiro, o STI (Speech Transmission Index) vem se tornando cada vez mais importante para a avaliação da inteligibilidade em lugares públicos. Muitas vezes, o STI já faz parte dos requerimentos do cliente à construtora de uma nova edificação. Se o STI não estiver acima de um valor mínimo no prédio concluído, a construtora terá que adotar medidas para melhorar a qualidade acústica do recinto, senão estará sujeita a pena contratual. Por isso, o STI já tem sido avaliado na fase de planejamento por softwares de simulação para conferir a viabilidade dos requerimentos. Após a construção, os resultados são verificados mediante medições “on site”. Este artigo pretende descrever os conceitos fundamentais do STI e apresenta a técnica e o processamento para obtê-lo. Alguns exemplos teóricos e práticos são apresentados e alguns pontos fracos do STI revelados.*

## 1. Introdução

Obter uma qualidade acústica satisfatória e uma boa inteligibilidade da voz humana em lugares onde é primordial que um locutor seja ouvido e entendido por todas as pessoas presentes (por exemplo auditórios, salas de aulas, estádios, igrejas, aeroportos, ferroviárias, rodoviárias, etc.) sempre foi um assunto complicado. A situação é pior no Brasil, devido à sua condição de país tropical, que reflete na maneira de construir. Para garantir uma boa ventilação nos dias quentes de verão, é comum deixar amplos vãos na alvenaria, pelos quais o ar, mas infelizmente também o ruído, circula entre o exterior e interior do prédio. Os onipresentes aparelhos de ar condicionado causam um nível constante de ruído de fundo que atrapalha

bastante o entendimento, como pôde ser experimentado no próprio auditório da faculdade de engenharia civil da Unicamp que sediou a segunda edição do simpósio da AcMus na qual este trabalho foi apresentado.

Como as janelas brasileiras geralmente não têm projeto de vedação contra o frio, elas são de feitiço simples e possuem molduras simples de alumínio para o deslizamento que também não vedam efetivamente o barulho oriundo das ruas.

Para agravar a situação, os níveis de ruído nas cidades brasileiras são notavelmente acima dos encontrados nos países mais ricos, pela falta de controle e pela má conservação de veículos e vias de circulação. Recentemente, os ônibus municipais começaram a ser equipados com ar condicionado, cujo compressor no teto é uma indesejável fonte adicional de ruído.

Finalmente, também existe um fator cultural: Shows ao vivo, comícios, cultos evangélicos, anúncios por carros com caixas de som enormes no teto e até conversa nos botequins e restaurantes, tudo parecendo precisar de um nível de som exageradamente alto.

Enfrentamos então dois agravantes no Brasil: de um lado, um elevado ruído de fundo, e do outro, edificações com pouco isolamento contra o ruído. Como terceiro agravante, tratamento acústico com absorvedores dentro dos prédios é coisa rara, devido aos custos, mas também como medida de proteção à higiene e à saúde pela facilidade que os materiais absorventes têm em acumular sujeira e mofo devido às altas temperaturas e taxas de umidade do ar.

Em lugares cuja função principal é de passar informações de forma oral de uma pessoa para um grupo de outras (escolas, universidades, auditórios), ou onde existe uma grande concentração de pessoas que, num caso de emergência, precisam receber avisos inteligíveis mesmo se houver pânico e gritaria (estádios, aeroportos, salas de show), faz sentido de prescrever uma inteligibilidade mínima em todos os pontos do público.

A maneira mais certa de avaliar a qualidade da transmissão de fala humana é de conduzir extensos testes de inteligibilidade com listas de palavras ou sílabas, envolvendo muitas pessoas espalhadas na área útil do lugar, e aplicar métodos de avaliação estatística. No entanto, esse procedimento é muito moroso e não permite a previsão da inteligibilidade ainda na fase de planejamento de uma edificação.

Dos múltiplos parâmetros acústicos objetivos que podem ser obtidos mediante simulação e medição da conjuntura acústica, o que mais se popularizou nas últimas três décadas e que finalmente foi introduzido em normas e regulamentos de vários países, é o STI (Speech Transmission Index). A razão disso é que ele leva em consideração tanto a reverberação e os ecos das salas quanto o ruído de fundo, sendo portanto sensível aos dois maiores inimigos da inteligibilidade. Além disso, ao contrário de outros parâmetros acústicos conhecidos, o resultado é um simples índice, de fácil interpretação, que ocupa valores entre 0 (fala completamente ininteligível) e 1 (ótima inteligibilidade).

## 2. História

Os dois idealizadores do STI, os holandeses Tammo Houtgast e Herman J.M. Steeneken, já se envolviam em testes de inteligibilidade no final dos anos 60. Na época, eles foram solicitados para pesquisar o alcance da transmissão de rádios VHF. Os testes, conduzidos pelo método tradicional de avaliação subjetiva, consumiram tempo considerável, o que incentivou os dois pesquisadores a desenvolver um método objetivo e bem mais rápido com base em sinais artificiais.

O resultado desse afã foi apresentado num artigo importante na revista *Acustica*, em 1971 (depois que o manuscrito foi refutado pela JASA!), seguido por inúmeras outras publicações da dupla, nas quais eles foram apresentando aperfeiçoamentos e verificações do método.

O STI, junto com o RASTI (Rapid STI), que é uma simplificação para cálculo mais rápido que hoje está caindo em desuso, culminou na norma IEC 268-16 em 1988. A segunda edição dessa norma entrou em vigor em 1998 e foi substituída em 2003 pela atual versão, a IEC 60268-16. Essa terceira edição introduziu duas versões do STI, uma direcionada para fala feminina e a outra para fala masculina, e leva em conta a redundância de resultados parciais em bandas de oitava adjacentes. Uma quarta edição está em fase de produção e deverá considerar outros aspectos práticos que ocorrem em situações típicas de medição. Também tentará resolver algumas ambigüidades e pontos mal-esclarecidos.

A descrição da obtenção do STI neste artigo se baseia na versão vigente, ou seja, na terceira edição de 2003 [1].

## 3. Conceito Básico

No início dos trabalhos, Steeneken e Houtgast [2-5] encetaram uma série de ensaios para avaliar propriedades estatísticas da fala humana e para descobrir quais dessas propriedades são sensíveis a mudanças em termos de inteligibilidade. Eles descobriram que a *intensidade*, ou seja, o quadrado do sinal temporal da voz, exibe componentes espectrais significativos na faixa de 0.5 a 25 Hz, com um valor máximo localizado na região entre 3 e 4 Hz.

Num ambiente com ruído de fundo ou reverberação notável, é claro que os mínimos da intensidade (as lacunas entre palavras e sílabas) são preenchidos, o que leva a uma aparente redução da profundidade da *modulação* da intensidade. No caso de ruído de fundo, esse preenchimento independe da frequência de modulação. Já no caso de reverberação, flutuações rápidas da intensidade sofrem mais nivelamento do que trechos com frequência de modulação lenta, onde a reverberação já decai bastante dentro do intervalo entre dois máximos da modulação.

Esses fatos levaram à concepção da *função de transferência de modulação*, ou *modulation transfer function (MTF)* em inglês. A MTF representa a transferência  $m(f)$  do envelope da intensidade da entrada à saída, dependendo da frequência  $f$  de modulação. O ruído age como um atenuador na MTF, enquanto que a reverberação exibe o efeito de um filtro passa-baixa.

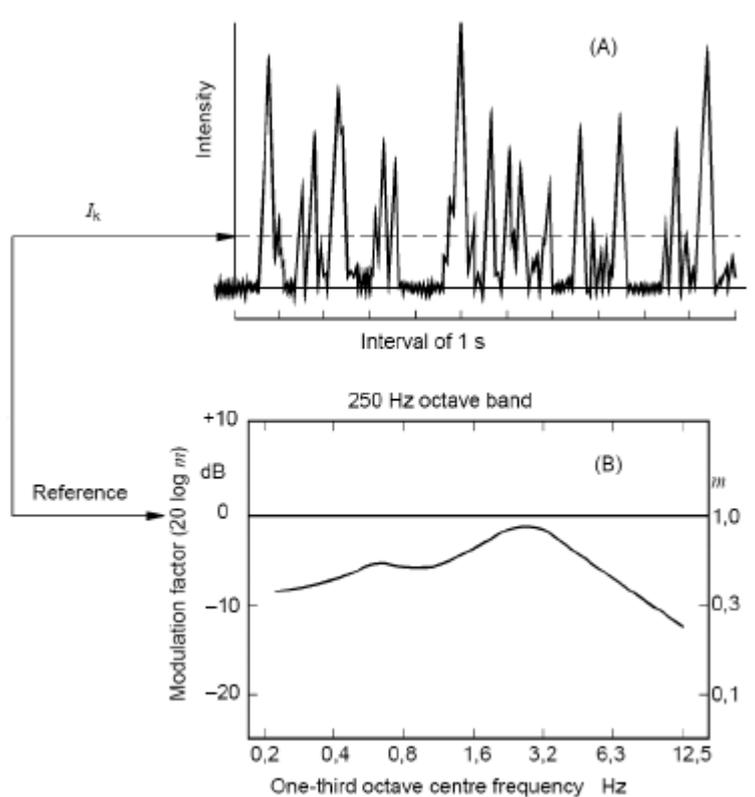


Fig. 1: Envelope típico da intensidade de fala e sua transformada de Fourier. De [1].

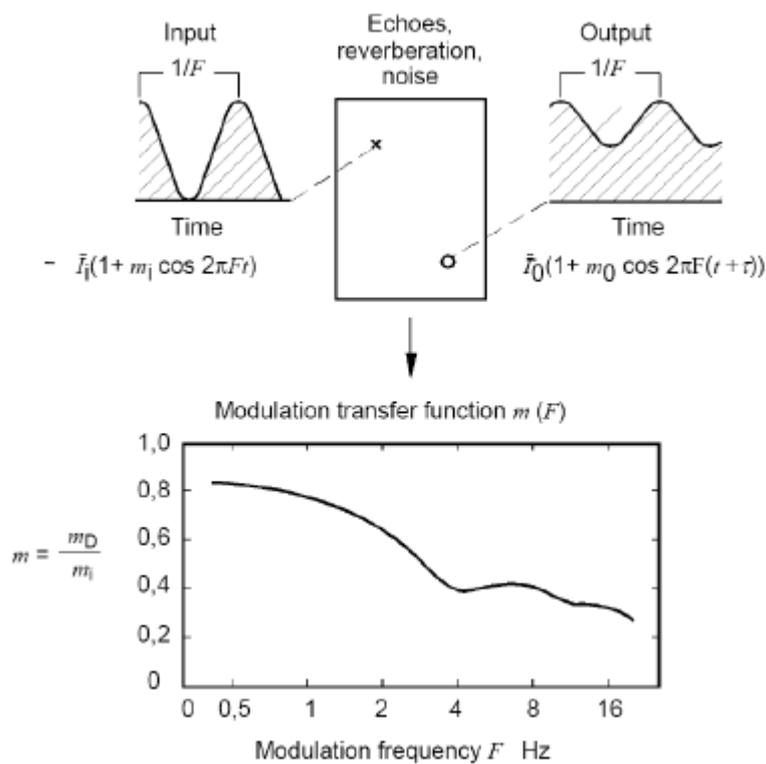


Fig. 2: Função da Transferência de Modulação, indicando influência de ruído, reverberação e ecos. De [1]

Para medir a MTF, a forma original do STI estipula a modulação senoidal da intensidade (ou seja, do quadrado) de um sinal de ruído. Isso corresponde à multiplicação do ruído com o termo  $v\{1+\cos(2\pi f_M t)\}$ , sendo  $f_M$  a frequência da modulação. Para poder ponderar várias faixas com importância variada para a transmissão da voz humana, o ruído é dissecado nas 7 oitavas entre 125 Hz e 8 kHz mediante passa-faixas. Cada um desses 7 ruídos filtrados, por sua vez, é modulado por 14 frequências, subindo em passos de terços de oitavas de 0.63 Hz até 12.5 Hz. Ao todo, são então 98 sinais diferentes, e cada um deles é reproduzido e capturado separadamente no ambiente sob investigação.

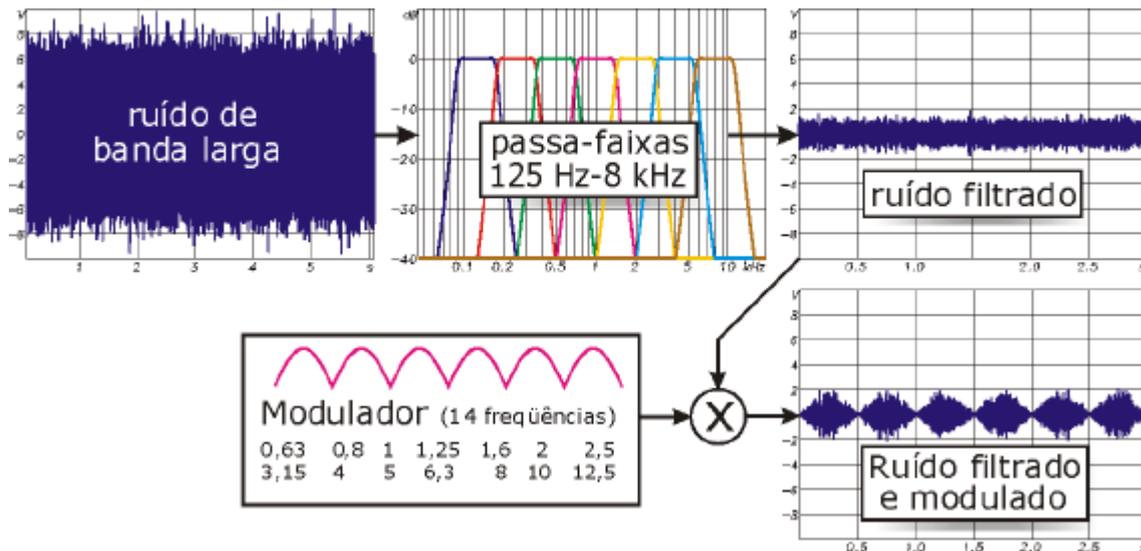


Fig. 3: Criação dos 98 sinais de teste para o método tradicional da obtenção do STI.  
Exemplo: banda de oitava  $k = 1$  kHz, frequência de modulação  $f = 2$  Hz.

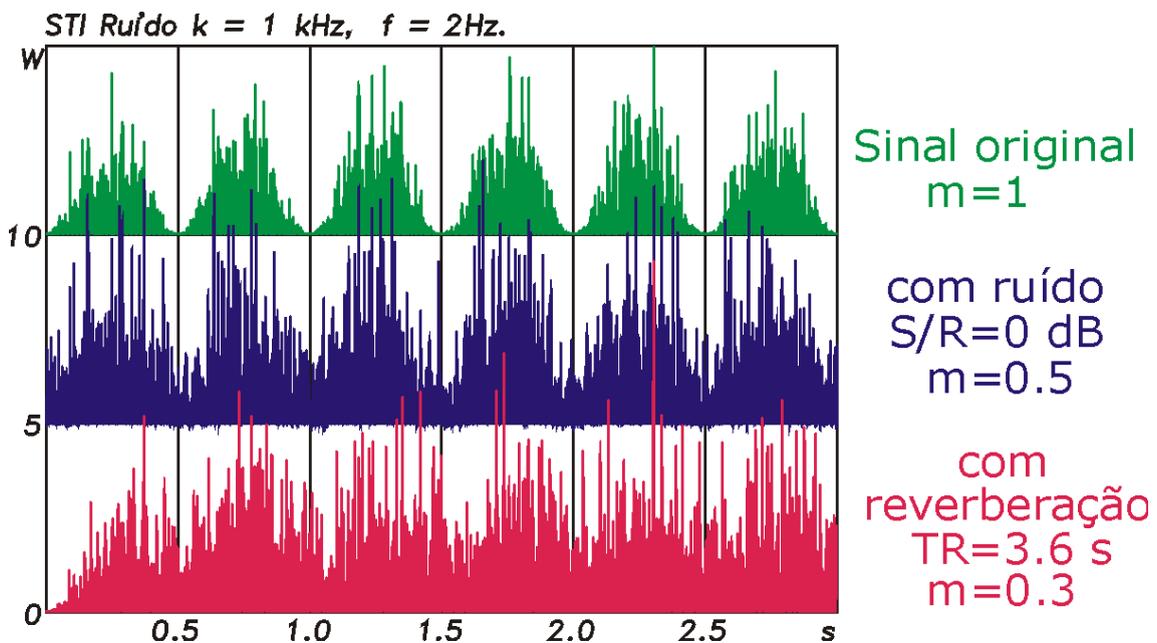


Fig. 4: Influência de ruído e da reverberação na transferência de modulação.  
Exemplo na banda de 1 kHz com frequência de modulação de 2 Hz.

### 3.1 Cálculo dos valores $m_{k,f}$

No ponto de recepção, o fator de redução da modulação  $m$  para cada um desses 98 sinais é avaliado da seguinte maneira:

- 1) Aplicar um filtro de oitava (que corresponde àquele utilizado na filtragem do ruído),
- 2) elevar o sinal filtrado ao quadrado,
- 3) aplicar ao sinal filtrado e elevado ao quadrado uma transformada discreta de Fourier de uma frequência só, isto é, a da modulação. Em outras palavras, multiplicar o intervalo analisado com o seno e o cosseno da frequência de modulação,
- 4) dividir o módulo do resultado da análise de Fourier, ou seja  $\sqrt{\text{Re}^2 + \text{Im}^2}$ , pela energia total, quer dizer, a soma da intensidade no intervalo analisado,
- 5) e finalmente multiplicar esse resultado com 2.

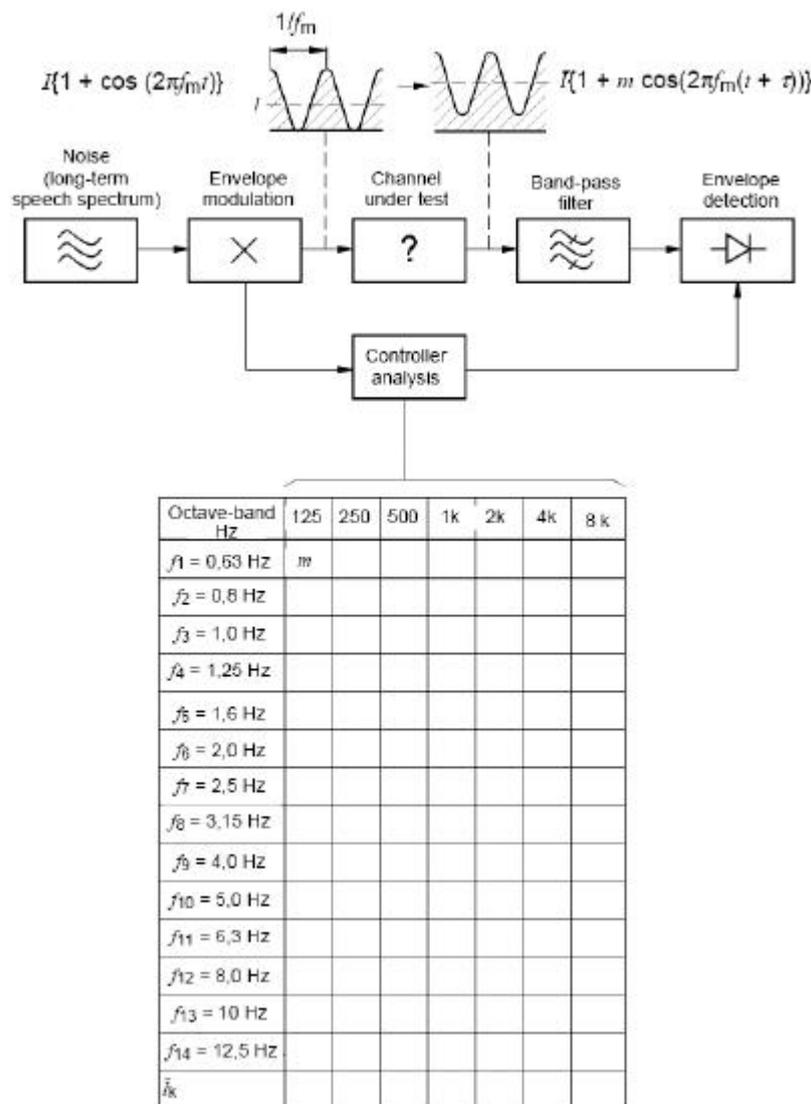


Fig. 5: Organização dos Resultados na matriz de MTFs. De [1]

Banda de oitava	-	k	1	2	3	4	5	6	7
Frequência central	Hz		125	250	500	1000	2000	4000	8000
Frequência de modulação	0,63	1	0,61492	0,63433	0,71096	0,74323	0,79465	0,93494	0,99339
	0,8	2	0,5475	0,56909	0,65291	0,6886	0,74626	0,91249	0,99067
	1	3	0,47329	0,49423	0,58963	0,6331	0,69634	0,89577	0,98651
	1,25	4	0,42455	0,40989	0,53317	0,57831	0,64925	0,86708	0,98322
	1,6	5	0,37591	0,3177	0,49855	0,551	0,60244	0,82768	0,97714
	2	6	0,32873	0,26033	0,47709	0,56036	0,56429	0,80212	0,97191
	2,5	7	0,28325	0,24494	0,43642	0,52633	0,53164	0,78004	0,96414
	3,15	8	0,23936	0,2524	0,40593	0,50771	0,50469	0,76227	0,95457
	4	9	0,20448	0,27341	0,41424	0,47877	0,48419	0,74187	0,94587
	5	10	0,22861	0,2322	0,38552	0,47717	0,46186	0,72289	0,93561
	6,3	11	0,12526	0,20446	0,38593	0,44943	0,46705	0,706	0,92367
	8	12	0,10567	0,26767	0,3414	0,44252	0,45744	0,88624	0,91187
	10	13	0,12026	0,1889	0,32927	0,4289	0,43267	0,84697	0,8951
	12,5	14	0,090121	0,21126	0,31575	0,39218	0,39427	0,81032	0,87234

Fig. 6: Matriz com valores  $m$  típicos de um ensaio real, com destaque do valor  $m$  para a combinação de banda de oitava  $k$  de 1 kHz e frequência de modulação  $f$  de 2 Hz.

Os 98 resultados diferentes são organizados na chamada matriz de MTF, uma tabela com as frequências de modulação agrupadas nas linhas e as frequências centrais do ruído filtrado nas colunas. Essa matriz serve como base para o cálculo do próprio STI, apresentado no capítulo 5. Mas vamos primeiro conhecer um método bem mais rápido para chegar à matriz dos valores  $m(f)$ :

#### 4. Avaliação rápida dos valores $m(f)$ mediante resposta impulsiva

O método tradicional, que consiste em tocar e analisar cada uma das 98 combinações de ruído modulado, é muito moroso, consumindo no mínimo 15 minutos. Isso o restringe aos casos nos quais meramente uma comunicação ponto a ponto esta sendo investigada (por exemplo, sistemas de comunicação com *headset*, como os utilizados pelos pilotos de aviões). Para a avaliação do STI em áreas com cobertura para muitas pessoas, é imprescindível medir o STI em vários pontos a fim de poder executar médias espaciais. Nesse caso, a duração de cada avaliação do STI completo é proibitiva. Uma primeira tentativa de diminuir o tempo total seria de tocar ruído simultâneo em todas as bandas. Como o primeiro passo no ponto de recepção é o de isolar cada banda com filtros de oitavas, é possível obter 7 valores  $m$  concomitantemente, contanto que a seletividade dos passa-faixas seja suficiente. Uma derivada do STI concebida especialmente para a avaliação da inteligibilidade de sistemas de sonorização, o STIPA (STI para PAs, PA = Public Address), vai a um passo além: Ela só utiliza seis bandas de ruído, das quais cada uma é modulada com *duas* frequências (relação 1:5). Cada uma das duas só chega a uma profundidade de modulação de  $m=0,55$ , ao contrário do  $m=1$  no método tradicional. Todas as bandas são tocadas simultaneamente, o que reduz o tempo de medição e avaliação a algo em torno de tipicamente 12 segundos. Porém, como o método só explora duas frequências de modulação em cada banda, ele é menos preciso do que o método tradicional e em alguns casos específicos pode errar bastante na predição da inteligibilidade.

Por isso, é sempre recomendável medir o STI completo. Em vez do ruído modulado reproduzido seqüencialmente, existe um método bem mais rápido para chegar a todos os valores  $m$  da matriz. Ele explora as *respostas impulsivas* (RIs) do ambiente pesquisado. Segundo a teoria, a RI entre dois pontos descreve todas as características lineares do

caminho de transmissão. De certa forma, todas as informações sobre reflexões e reverberação devem então estar embutidas nela. De fato, a avaliação dos valores de transferência  $\mathbf{m}$  a partir de RIs funciona de forma muito parecida à dos ruídos modulados:

- 1) Aplicar um banco de 7 filtros de oitava, de 125 Hz até 8 kHz, à RI de banda larga. Os filtros devem cumprir a norma IEC 1260. Filtros passa-faixa de ordem 6 a 8, do tipo Butterworth, ou melhor Chebychev com baixo *ripple*, servem para o propósito.
- 2) Elevar os sete sinais resultantes ao quadrado,
- 3) Aplicar a cada um dos 7 sinais filtrados e elevados ao quadrado uma transformada discreta de Fourier com as 14 frequências exatas (entre 0.63 Hz e 12.5 Hz) de modulação estipuladas pela norma do STI. Mais uma vez, isso significa multiplicar o intervalo analisado com o seno e o coseno de cada frequência de modulação, o que resulta em 14 partes reais e imaginárias para cada banda de oitava. Para evitar erros de vazamento, só um número inteiro de períodos deverá ser analisado para cada frequência de modulação. Como a frequência mais baixa de modulação é de 0.63 Hz, isso também significa que as RIs medidas devem acolher no mínimo um período inteiro dessa frequência, ou seja, o comprimento da RI não deve ser inferior a 1.6 segundos. Caso contrário, erros no cálculo dos valores  $\mathbf{m}$  para as frequências baixas de modulação são inevitáveis. Em ambientes com muita reverberação, é claro, o comprimento da RI deve ser maior ainda para não cortar partes significantes da cauda reverberante, o que levaria a uma superestimação do STI.

Alternativamente, uma FFT pode ser aplicada sobre todo o intervalo e os valores nas frequências exatas de modulação podem ser obtidos por interpolação entre os valores vizinhos do espectro FFT. Mas como são poucas as frequências avaliadas, a FFT não é o cálculo mais eficiente nesse caso. Além disso, a interpolação e a falta da possibilidade de restringir a análise a intervalos inteiros das frequências exatas de modulação tornam esse método menos preciso.

- 4) Finalmente, dividir o módulo dos 14 resultados da análise de Fourier em cada banda, ou seja  $\sqrt{\text{Re}^2 + \text{Im}^2}$ , pela energia total dessa banda (a parte DC), o que resulta diretamente nos 14 valores  $\mathbf{m}$  de cada banda.

Tudo isso significa que uma RI só, que na maioria das vezes é levantada de qualquer maneira para avaliar outros parâmetros acústicos (sobretudo os tempos de reverberação), é suficiente para chegar rapidamente aos 98 valores  $\mathbf{m}$  que compõem a matriz das MTF.

Porém, existe uma grande diferença em relação ao método tradicional. O intuito na captação de RIs normalmente é de diminuir ao máximo a influência do ruído de fundo, para não atrapalhar o cálculo dos tempos de reverberação e demais parâmetros acústicos. Para chegar a essa meta, o sinal de excitação normalmente é reproduzido com volume alto, e médias síncronas podem ser efetuadas para melhorar ainda mais a relação S/R. Isso significa que a matriz dos valores  $\mathbf{m}$  e o próprio STI obtidos a partir da RI são praticamente isentos da influência do ruído de fundo. Para incluir ele devidamente neste *noiseless STI*, existem dois métodos. O primeiro é de utilizar como sinal de excitação um ruído pseudo-aleatório (por

exemplo, MLS) e pré-filtrar ele com um filtro de fala (speech filter) a fim de obter uma distribuição espectral que corresponde à da fala humana. Esse sinal deve ser tocado com o mesmo valor RMS ao que seria emitido por um locutor numa situação típica de anúncio no ambiente pesquisado. Médias não devem ser feitas. O resultado será que o ruído de fundo se espalhará com a proporção certa na RI e conseqüentemente terá reflexo correto na diminuição do STI.

Ainda assim, esse método é pouco prático e pouco recomendável. As RIs captadas dessa maneira normalmente não poderão mais ser utilizadas para a avaliação dos outros parâmetros acústicos (a não ser que o ruído de fundo seja muito baixo, mas neste caso, o procedimento descrito seria desnecessário de qualquer maneira). Uma solução muito melhor é medir separadamente o nível **R** do ruído de fundo e também o nível **S+R** de uma pessoa falando com volume típico no ambiente examinado. Ambos os níveis são levantados nas bandas de oitavas entre 125 Hz e 8 kHz. O nível do orador é calculado através do L10 ou L12 (nível de pressão sonora que está sendo ultrapassado em 10% ou 12% do tempo, respectivamente), ou através do valor RMS da fala, desconsiderando as folgas entre as sílabas. Quando a medição do nível do orador no lugar examinado também inclui o ruído **R** de fundo, esse tem que ser excluído para chegar-se ao valor real do nível **S** do orador:

$$S = 10 \cdot \log(10^{S+R/10} - 10^{R/10})$$

A relação S/R servirá para manipular todos os 14 valores **m** de uma banda de oitava *k* por um termo de correção que será apresentado no capítulo 6.

Esse procedimento, além de ser mais prático e de execução mais fácil, traz uma grande vantagem: A possibilidade de discernir o efeito do ruído de fundo e o da acústica da sala, prevendo a influência no STI quando só um dos dois varia.

## 5. Filtragem rápida das respostas impulsivas

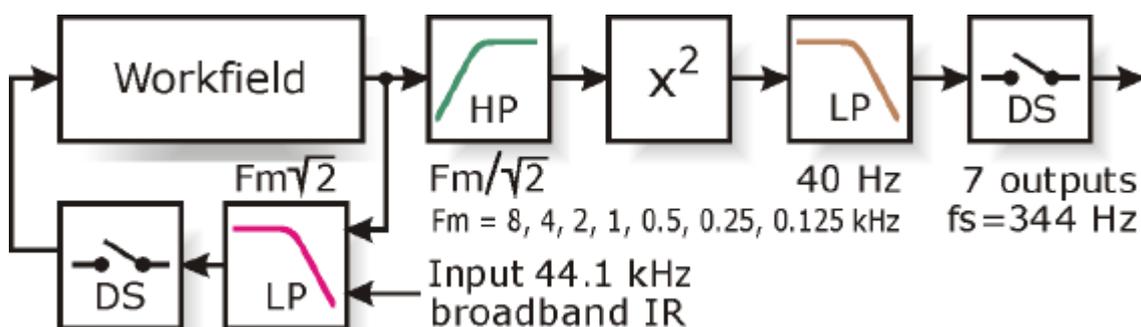
Embora o cálculo descrito acima possa ser efetuado em tempo razoável em computadores modernos com velocidade de processamento na casa dos GHz, é possível apressar ele ainda mais a fim de obter os resultados finais quase instantaneamente, mesmo analisando RIs extensas ou de vários pontos de recepção em conjunto.

Uma maneira eficiente de reduzir o custo computacional é realizando reduções sucessivas da taxa de amostragem (*downsampling*) durante a filtragem em bandas de oitava. Após a aplicação do passa-baixa que define o limite superior de cada banda de oitava, a taxa de amostragem pode ser diminuída em passos de 2 (quer dizer, cada segundo valor do sinal é simplesmente descartado), observando um limite razoável de *aliasing* (componentes espelhados) que possa entrar no resultado. Um critério razoável é garantir que a nova taxa de amostragem fique no mínimo 3 a 4 vezes acima da frequência de corte do passa-baixa.

Após a elevação ao quadrado, outro *downsampling* substancial pode ser efetuado, uma vez que a análise dos componentes espectrais das RIs filtradas e quadradas se estende meramente até 12.5 Hz. Portanto, a frequência final de amostragem pode ser reduzida para algo em torno de 100 a 300 Hz. Nesse caso, o cálculo das transformadas discretas de Fourier é concluído

tão rapidamente que a contribuição ao tempo total do cálculo do STI é negligenciável. Levando isso em consideração, um passa-baixa de ordem baixa (2 a 4) pode ser utilizado para filtrar os sinais elevados ao quadrado antes de reduzir a taxa de amostragem deles ao valor final. Manter ela 10 ou 20 vezes acima da maior frequência considerada (12.5 Hz) garante que o efeito de componentes aliasing seja desprezível, não obstante a inclinação branda de um passa-baixa de ordem tão baixa.

Essas considerações levam ao processamento da RI de banda larga apresentado na Fig. 7. Ele procede da banda de oitava mais alta (8 kHz) até a banda mais baixa (125 Hz). Em cada ciclo, a RI é primeiramente tratada com um passa-baixa com frequência de corte de  $\sqrt{2} f_m$  ( $f_m$  = frequência nominal da banda). Em seguida, a taxa de amostragem é reduzida para a metade e o resultado é guardado num campo temporário (“*workfield*” na Fig. 7). Depois disso, um passa-alta com frequência de corte de  $f_m/\sqrt{2}$  é aplicado. A saída desse filtro é elevada ao quadrado e filtrada com outro passa-baixa de ordem 4 e frequência de corte de 40 Hz. Finalmente, a taxa de amostragem desse sinal é diminuída para o valor final. Isso o deixa pronto para ser submetido à análise de Fourier, resultando nos 98 valores  $m_{k,f}$ .



**Fig. 7: Processamento rápido com redução sucessiva da taxa de amostragem para estimar a RI filtrada quadrática nas 7 bandas de oitava.**

O ganho computacional é considerável: Num PC equipado com P4 de 2.4 GHz, o cálculo do STI a partir de uma RI de 6 segundos (frequência de amostragem: 44.1 kHz) é completado em menos de 200 ms.

## 6. Cálculo do STI a partir dos valores $m$

Os 98 valores  $m_{k,f}$  constituem a base de cálculo do STI. Mas eles ainda sofrerão algumas correções e transformações antes de entrar na equação final.

A primeira e mais importante correção, já anunciada no capítulo 4, é a que introduz a relação S/R, contanto que os valores  $m$  tenham sido derivados devidamente de uma RI com influência desprezível do ruído:

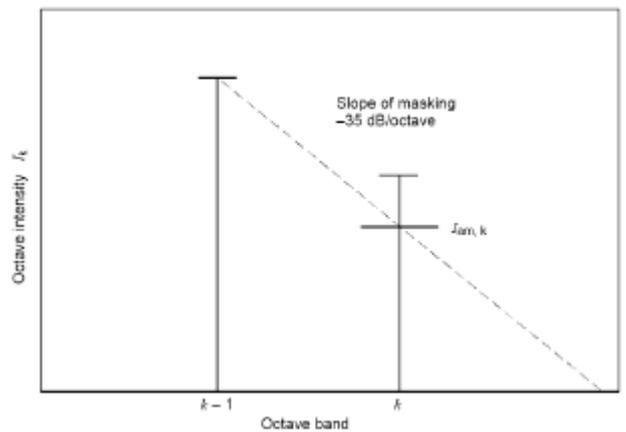
$$m_{COR\_SNR} = m \cdot \frac{1}{1 + 10^{\frac{-S/R}{10}}}$$

Desde a terceira edição da norma IEC 60268-16, existem mais duas correções em bandas de oitavas que dizem respeito ao *mascamamento* em bandas adjacentes, e ao *nível absoluto* da

fala do locutor, respectivamente. Ambas as correções são modeladas como um ruído adicional que piora a transferência de modulação, ou seja, baixa os valores  $\mathbf{m}$ .

### 6.1 Consideração do mascaramento e do limiar de audição

O mascaramento diz respeito ao efeito psicoacústico de sobreposição de um som forte sobre um som fraco numa frequência vizinha. No caso do cálculo do STI, que trabalha com bandas de oitava, o mascaramento terá efeito notável se uma banda possuir um nível  $\mathbf{S}$  da fala bem mais baixo do que nas bandas adjacentes (o que pode acontecer em sistemas eletro-acústicos mal equalizados, por exemplo). Para simplificar o cálculo, só a banda inferior é considerada. O mascaramento dessa banda  $k-1$  sobre a banda  $k$  é modelada como uma linha que decai com uma declinação fixa em dB/oitava:



**Fig. 8: Efeito de mascaramento da banda  $k-1$  sobre a banda superior  $k$  para um nível entre 55 e 65 dB do sinal do locutor na banda  $k-1$ . De [1].**

Para complicar um pouco as coisas, a inclinação dessa reta depende do nível  $\mathbf{S}_{k-1}$  na banda  $k-1$ . Na norma, essa dependência é tabelada em passos discretos de 10 dB:

Octave level dB	46-55	56-65	66-75	76-85	86-95	>95
Slope of masking	-40	-35	-25	-20	-15	-10
Auditory masking factor	0,000100	0,000316	0,003162	0,010000	0,031622	0,100000

**Tabela 1: Inclinação do efeito de mascaramento dependendo do nível  $\mathbf{S}$  e o correspondente fator  $\mathbf{amf}$  (auditory masking factor) utilizado no cálculo. De [1].**

Felizmente, a inclinação pelo menos não depende da frequência. A influência do mascaramento é modelada como uma *intensidade* de som que interfere na banda  $k$  e entrará no termo de correção dos valores  $\mathbf{m}$ :

$$I_{am,k} = 10^{\frac{S_{k-1}}{10}} \cdot amf$$

$I_{am,k}$  é o mascaramento exercido na banda  $k$ ,  $S_{k-1}$  é o nível do mascarador na banda inferior  $k-1$ , e  $\mathbf{amf}$  é o fator de mascaramento dependendo desse nível  $S_{k-1}$  e determinado com a ajuda da Tabela 1.

Quanto mais o nível de fala cair e se aproximar do limiar de audição, maior terá que ser a relação S/R para manter a mesma inteligibilidade. Para dar conta desse efeito, um terceiro componente, o *limiar absoluto de recepção* (absolute reception threshold) entra no termo de correção dos valores  $\mathbf{m}$ . Esse limiar depende da frequência. Os valores em dB podem ser encontrados na Tabela 2. Para o cálculo, eles são transformados para intensidades:

$$I_{rs,k} = 10^{\frac{L_{rs,k}}{10}}$$

Finalmente, com a ajuda da própria intensidade  $I_k$  da fala na banda  $k$ ,

$$I_k = 10^{\frac{S_k}{10}}$$

o termo de correção  $mcor_k$ , incluindo também a influência da relação S/R, se torna:

$$mcor_k = \frac{1}{1 + 10^{\frac{-S/R}{10}}} \cdot \frac{I_k}{I_k + I_{am,k} + I_{rs,k}}$$

Todos os 14 valores  $\mathbf{m}_{k,f}$  de determinada banda  $k$  agora podem ser multiplicados com esse termo de correção:

$$m'_{k,f} = m_{k,f} \cdot mcor_k$$

Resta ressaltar que as duas correções relativas ao limiar absoluto de audição e ao mascaramento só têm efeito significativo se o nível de fala for baixo e divergir muito em bandas adjacentes, respectivamente.

#### 6.4 Cálculo dos índices de transmissão e médias finais

Após esse ajuste dos valores  $\mathbf{m}$ , o STI pode ser calculado. Como primeiro passo, cada valor  $\mathbf{m}$  é transformado em uma relação sinal/ruído (signal to noise ratio, SNR) aparente:

$$SNR_{k,f} = 10 \cdot \log \frac{m'_{k,f}}{1 - m'_{k,f}} \text{ dB}$$

Se a redução da transferência de modulação fosse exclusivamente ocasionada pelo ruído de fundo, esse valor SNR de fato apresentaria a relação sinal/ruído real.

Todos os 98 valores SNR são ceifados a uma faixa de  $\pm 15$  dB. Quer dizer, um valor muito bom que ultrapassa 15 dB é fixado em 15 dB, e um valor péssimo que cai abaixo de -15 dB é ajustado em -15 dB.

Em seguida, cada um dos valores  $SNR_{k,f}$  engastado na faixa de  $\pm 15$  dB é transformado no chamado *índice de transmissão* (transmission index, TI):

$$TI_{k,f} = \frac{SNR_{k,f} + 15 \text{ dB}}{30 \text{ dB}}$$

Devido à ceifa anterior dos valores  $SNR_{k,f}$ , os  $TI_{k,f}$  só podem tomar valores entre 0 e 1. Portanto, eles já se assemelham com o STI final, e de fato este é calculado em duas etapas

como média sobre todos os 98 índices de transmissão. A primeira etapa é uma simples média aritmética dos 14 índices dentro de cada banda de oitava:

$$MTI_k = \frac{1}{14} \sum_{f=1}^{14} TI_{k,f}$$

Este *índice de transmissão de modulação* (Modulation transfer index,  $MTI_k$ ) por oitava então dá igual peso a todas as 14 frequências de modulação na faixa de 0.63 Hz a 12.5 Hz considerada pelo STI. Mas existem propostas para ponderar mais a faixa entre 2 e 6.3 Hz, que tem maior importância para a inteligibilidade [9].

Finalmente, os 7 valores  $MTI_k$  formam o STI através da seguinte soma ponderada:

$$STI_r = \sum_{n=1}^7 a_n \cdot MTI_n - \sum_{n=1}^6 b_n \sqrt{MTI_n \cdot MTI_{n+1}}$$

A parte essencial é a primeira soma, que calcula a média dos  $MTI_k$ , mas nesta etapa não dando igual importância a todos os componentes do somatório. A segunda soma é um termo de redundância só introduzido a partir da terceira versão da norma IEC 60268-16. Ela considera que um resultado bom em uma banda pode parcialmente compensar um resultado pior numa banda adjacente. Os fatores de ponderação  $a$  e  $\beta$ , específicos para cada banda de oitava, dependem do sexo do locutor, o que é outra novidade da terceira edição da norma:

Octave band Hz		125	250	500	1 k	2 k	4 k	8 k
Males	$\alpha$	0,085	0,127	0,230	0,233	0,309	0,224	0,173
	$\beta$	0,085	0,078	0,065	0,011	0,047	0,095	–
Females	$\alpha$	–	0,117	0,223	0,216	0,328	0,250	0,194
	$\beta$	–	0,099	0,066	0,062	0,025	0,076	–
Absolute reception threshold dB	$L_{rs,k}$	46	27	12	6,5	7,5	8	12

**Tabela 2: Fatores de ponderação específicos para homens e mulheres e limiar absoluto de recepção, tudo em bandas de oitava. De [1].**

A banda mais baixa, 125 Hz, nem é considerada para mulheres, cuja voz normalmente tem fundamental só a partir de 150 Hz. Para os dois sexos, pesos maiores são dados nas oitavas entre 500 Hz e 4 Hz.

O pequeno índice  $r$  na sigla  $STI_r$  significa “revised” (revisado) e indica que as devidas correções do fator  $m$  foram aplicadas. Indica igualmente que o termo de redundância foi utilizado na soma final, junto com os novos coeficientes separados para mulher e homem.

## 7. Avaliação do valor do STI

Valor do STI	Avaliação segundo a IEC 60268-16
0.75 - 1	Excelente
0.6 - 0.75	Bom
0.45 - 0.6	adequado
0.3 - 0.45	Fraco
0 - 0.3	péssimo

Tabela 3: Relacionamento do valor STI com a inteligibilidade predita.

## 8. Dedução de outras medidas de inteligibilidade

Duas outras medidas correntes para descrever a inteligibilidade podem ser extraídas do STI mediante equações simples. Trata-se do valor da *Common Intelligibility Scale*, CIS:

$$CIS = 1 + \log(STI), 0 \leq CIS \leq 1$$

e do índice de articulação *%AlCons*, que é calculado através de uma equação empírica segundo Farell-Becker:

$$\% AlCons = 170.5405 \cdot e^{-5.419 \cdot STI}, 0 \leq \% AlCons \leq 100$$

## 9. Mecanismos básicos que influenciam no STI

Neste capítulo, iremos conhecer alguns casos teóricos e os efeitos que eles exercem sobre o STI. Para simplificar o panorama, as correções que dizem respeito ao mascaramento e ao limiar absoluto foram deixados fora do cálculo.

### 9.1 Influência de ruído

Ruído estacionário exerce uma influência na MTF que independe da frequência de modulação. A influência pode ser diretamente calculada através da equação já conhecida:

$$m_{SNR} = \frac{1}{1 + 10^{\frac{-S/R}{10}}}$$

Uma relação S/R de -15 dB ou pior em todas as bandas significa que o STI fica no patamar do “pior possível”, ou seja, 0. A partir desse limiar, cada aumento de 3 dB na relação S/R faz com que o STI aumente em 0.1, até finalmente chegar ao valor 1 (indicando inteligibilidade perfeita) quando a relação S/R atinge e ultrapassa 15 dB.

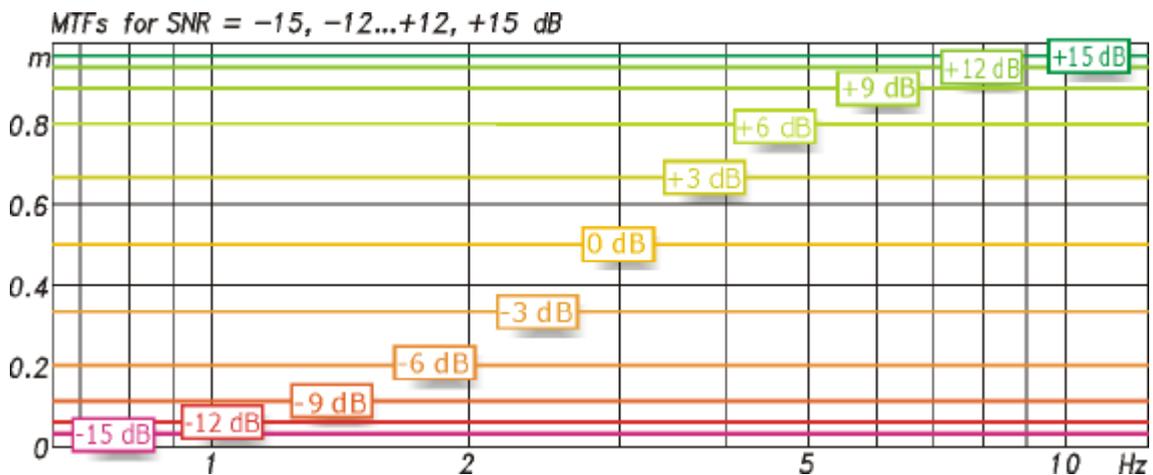


Fig. 9: Influência da relação S/R de banda larga na MTF

SNR [dB]	-15	-12	-9	-6	-3	0	+3	+6	+9	+12	+15
STI	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

Tabela 4: Relação entre SNR de banda larga e o STI.

Esse relacionamento se deve à projeção linear da relação S/R à faixa dos índices de transmissão entre 0 e 1:  $TI = (S/R + 30 \text{ dB}) / 15 \text{ dB}$ .

Outro fato interessante é que uma relação S/R de 0 dB em todas as bandas de oitava resulta num STI de exatamente 0.5, perto da divisa entre “adequado” e “fraco”.

## 9.2 Influência de reverberação

Como já mencionado, reverberação surte um efeito passa-baixa na MTF. Quer dizer, afeta bastante as altas frequências de modulação, enquanto tem efeito mais brando nas baixas. Contanto que o decaimento seja estritamente exponencial, indicando condições perfeitas de campo difuso, a influência de reverberação na MTF pode ser expressa pela seguinte equação:

$$m_{TR}(f) = \frac{1}{\sqrt{1 + \left(\frac{2p \cdot f \cdot TR}{13.8}\right)^2}}$$

Com um tempo de reverberação  $TR$  de 1 segundo em todas as bandas, o STI atingiria meramente um valor de 0.59, o que significa que ele já sairia da categoria “boa” para a “adequada”. Com 2 segundos de  $TR$ , o STI cairia para 0.44, o que já corresponde à categoria “fraco”. Na praxe, os tempos de reverberação raramente são iguais em todas as bandas de oitava. Tendem ser mais altos nas bandas baixas e mais bem comportados nas médias e nas altas. As bandas com maior ponderação no cálculo do STI são as de 500 Hz até 4 kHz.

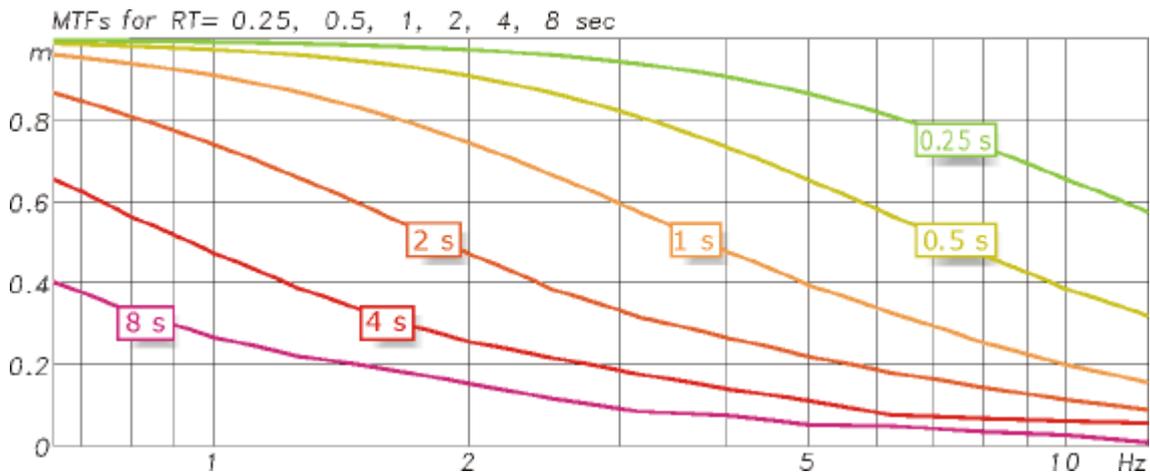


Fig. 10: Influência de reverberação de banda larga na MTF

RT [s]	8	4	2	1	0.5	0.25
STI	0.19	0.31	0.44	0.59	0.74	0.85

Tabela 5: Relação entre reverberação de banda larga e o STI

### 9.3 Influência de um eco

Ecos causam um efeito “filtro de pente” na MTF. Isso se deve à interferência construtiva e destrutiva da adição com o som direto causando “batimento”. A influência de um eco de banda larga com nível variável também pode ser calculada por uma equação:

$$m_{ECO}(f) = \frac{\sqrt{1 + 2 \cdot I_E \cdot \cos(2\pi \cdot f \cdot T_E) + I_E^2}}{1 + I_E} \quad \text{com } I_E = 10^{\frac{L_E}{10}}$$

O exemplo escolhido aqui trata com um eco de banda larga que chega 100 ms após o som direto. A Tabela 6 indica que mesmo se chegar com nível igual ao do som direto, o STI sofrerá queda para apenas 0.68, no meio da faixa de avaliação “bom”.

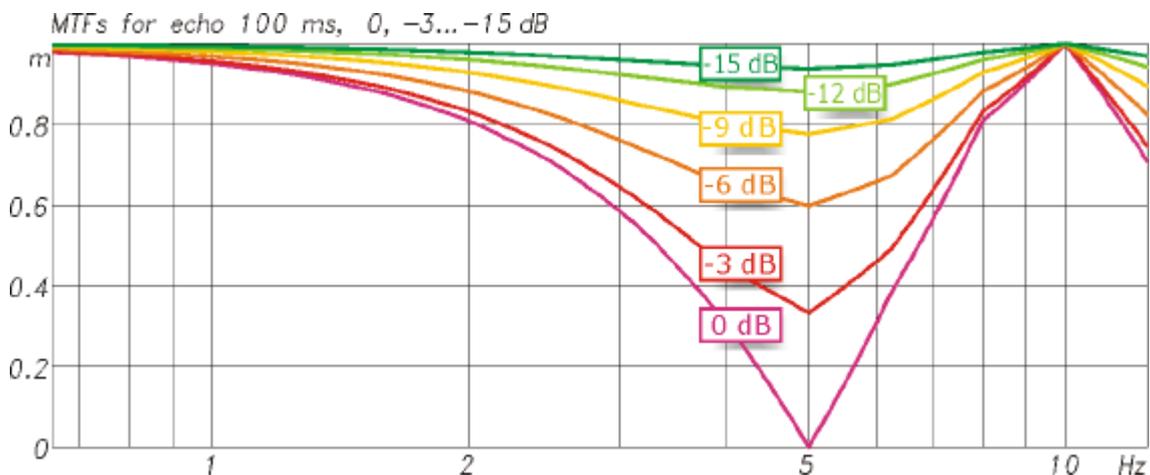


Fig. 11: Influência do nível de um eco de banda larga de 100 ms de atraso na MTF

$L_E$ [dB]	0	-3	-6	-9	-12	-15
STI	0.68	0.74	0.80	0.87	0.93	0.97

**Tabela 6: Relação entre nível de um eco de 100 ms de atraso e o STI**

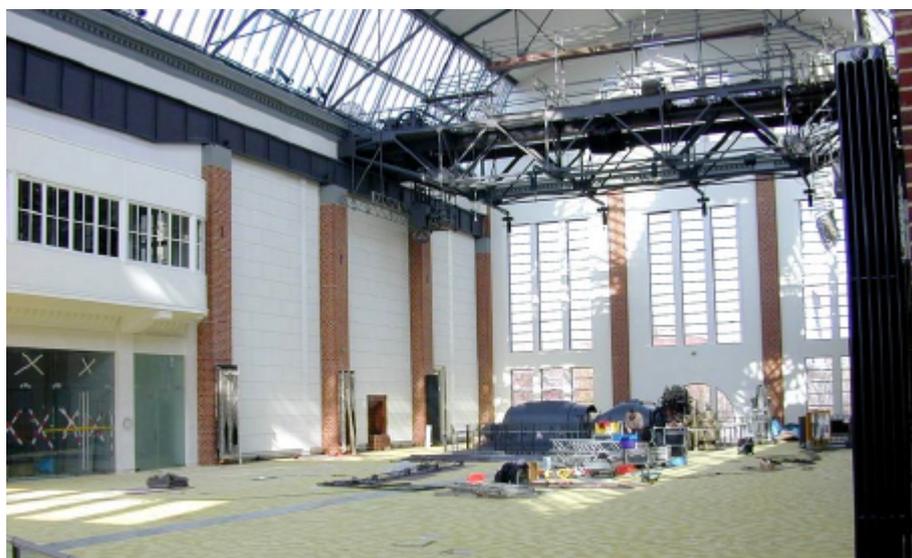
No entanto, esse tipo de eco prejudica bastante o entendimento, enquanto o STI não acusa suficientemente a perda de inteligibilidade. Essa insensibilidade a ecos tardios e nocivos à inteligibilidade é um ponto fraco do STI, que será corroborado com um exemplo prático no capítulo 10.3.

## 10. Algumas medições reais do STI

Vamos agora pôr as mãos na massa e conhecer algumas medições reais para nos familiarizar com o STI. Mais uma vez, as correções referentes ao mascaramento e ao limiar de audição não foram aplicadas, mas de qualquer maneira, elas surtem pouquíssimo efeito (uma piora de cerca 0.02) no resultado final.

### 10.1 Centro cultural em Düsseldorf

O primeiro exemplo é uma antiga central de energia na cidade de Düsseldorf, Alemanha, que foi transformada num centro cultural pela prefeitura.



**Fig. 12: Antiga Central de energia em Düsseldorf antes do tratamento acústico.**

As medições foram feitas com as obras de alvenaria já concluídas, mas ainda sem o revestimento de material absorvente. Os ensaios tiveram exatamente o propósito de prever qual a área que devia ser coberta com material absorvente para chegar a tempos de reverberação e STI razoáveis e compatíveis com o uso multi-funcional da sala.

A medição revelou que o tempo de reverberação da sala vazia era de 3.5 segundos. O STI, com um valor de 0.42, também indica condições complicadas. Ainda assim, uma auralização (convolução da RI com fala seca) mostrou que apesar da reverberação fortíssima,

ainda foi possível entender o locutor (que fala de maneira bastante pausada na gravação utilizada). O motivo disso é que o som direto, junto com uma primeira reflexão benéfica ocorrendo 10 ms depois, sobrepuja em 10 dB a reverberação (que demonstra um decaimento quase idealmente exponencial) que se estabelece depois. Isso faz toda a diferença, pois o nosso sentido de audição tem capacidade excelente de discernir a primeira chegada de som dos componentes atrasados.

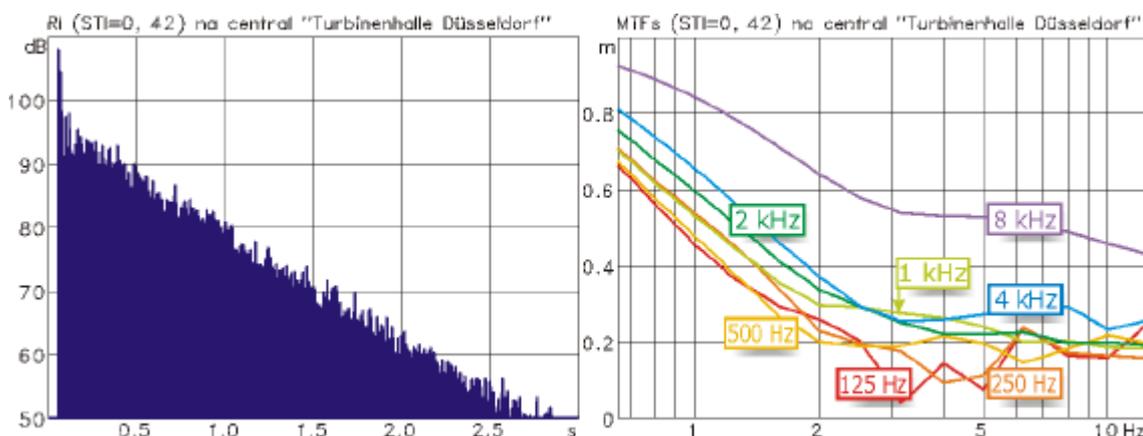


Fig. 13: Resposta impulsiva e MTFs na antiga central de energia.

## 10.2 Estádio Olympia em Berlim

O segundo exemplo é o estádio olímpico em Berlim que foi um dos utilizados na copa do mundo de 2006, inclusive para o jogo final entre França e Itália. No ano anterior, foi realizada uma reforma abrangente, incluindo um novo sistema de sonorização, composto por *line-arrays*, como de costume atualmente.

O primeiro passo em projetos desse tamanho normalmente é utilizar um software de predição da distribuição do som. Com os dados de direcionalidade e potência máxima dos alto-falantes e o modelo do recinto a ser sonorizado, esses softwares são capazes de calcular mapas com níveis na área de cobertura e até prever o STI (com base no som direto, um número reduzido de primeiras reflexões e na reverberação calculada com meios estatísticos). Executar essa simulação é um passo indispensável para verificar a adequação dos alto-falantes escolhidos. Nesse caso, a análise também foi exigida na licitação.

Após a instalação do sistema de sonorização, coube verificar se os valores de STI e o nível máximo preditos foram cumpridos. A uniformidade da resposta em frequência (pré-requisito:  $\pm 3$  dB na faixa 70 Hz – 15 kHz) também foi investigada. Para verificar essas metas de forma representativa, nada menos que 200 medições de RIs foram executadas em pontos espalhados sobre toda a área do público no estádio vazio. Em termos de acústica de sala, essa situação apresenta o pior caso, com a maior reverberação possível. Para levar o efeito absorvedor do público em consideração, um “bônus” conservador de 0.05 (calculado a partir da reverberação medida sem público e estimada com público) foi concedido ao valor final do STI.



**Fig. 14: Estádio “Olympia” em Berlin. De [8].**

Antes disso, a relação S/R foi modelada com base no nível **S** máximo alcançado pelo sistema de sonorização, predito pela simulação e verificado posteriormente com ruído rosa. Junto com um ruído **R** médio de 92 dB(A) estipulado para o público - em situações extremas, como a exaltação após um gol, esse pode chegar até 110 dB(A) -, chegou-se a um valor de 13 dB para a relação S/R, correspondendo a um fator de redução de 0.95 para todos os valores **m** da matriz das MTFs.

Para a amostra de RI apresentada na Fig. 16, essa correção S/R diminuiu o STI de 0.58 para 0.55, enquanto o bônus acrescentado resultou num valor final de 0.6, na fronteira entre as categorias “adequado” e “bom”. O pré-requisito estipulado pelo mandante da obra era um STI mínimo de 0.5 a ser cumprido em no mínimo 90% dos assentos no estádio lotado.

A reverberação no estádio vazio é enorme: 5 segundos na faixa entre 500 Hz e 2kHz. Mesmo assim, com as devidas correções, o STI dá “luz verde” neste exemplo. Mais importante: A convolução da RI com fala seca revela que tudo pode ser entendido sem maiores problemas, não obstante a cauda reverberante maciça que acompanha cada sílaba. A razão disso mais uma vez é a boa relação do som direto sobre a reverberação difusa. Como pode ser visto na Fig. 16, o som direto chega aos ouvintes do público com nível quase 20 dB acima da cauda reverberante. Isso é essencialmente fruto do uso de line-arrays, que agem como fonte de linha, emitindo ondas quase cilíndricas. Ao contrário de fontes pontuais, fontes de linha têm perda de somente 3 dB com cada dobro da distância, reforçando o som direto particularmente nas áreas mais afastadas dos alto-falantes.

Muitas tarefas de sonorização contam com esse problema: Nada ou pouca coisa pode ser feita para mudar uma acústica difícil que um recinto possa ter. A solução então é otimizar a radiação dos alto-falantes, buscando cobertura homogênea só na área do público com emissão de ondas coerentes.

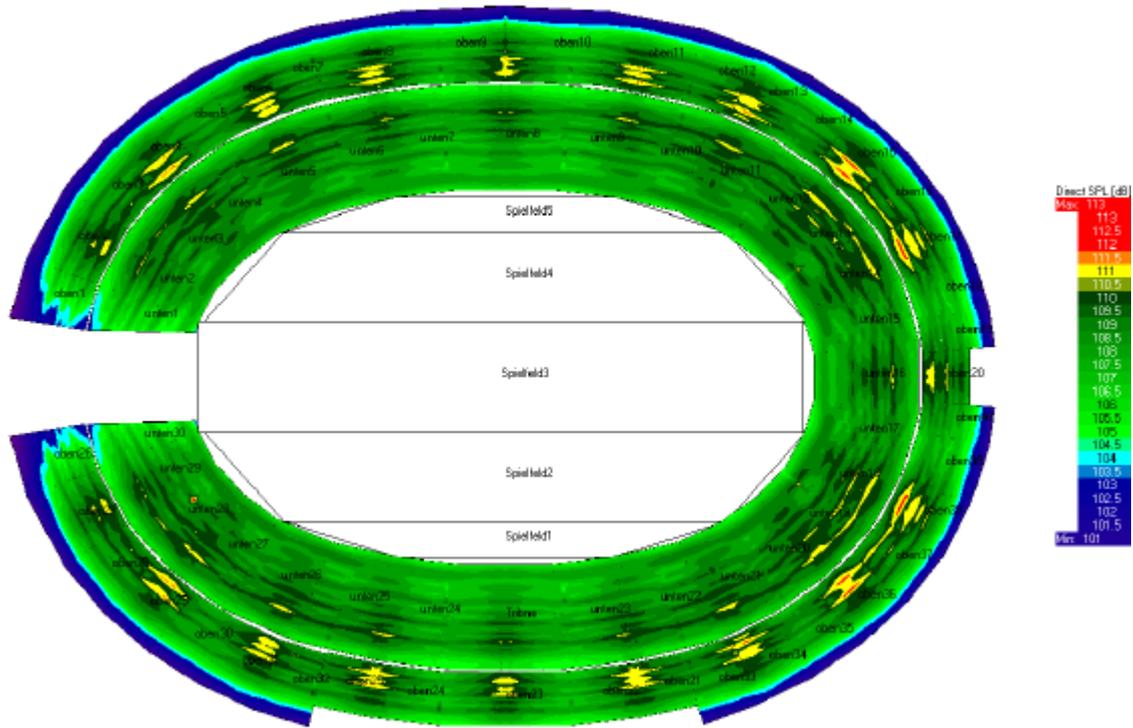


Fig. 15: Simulação da distribuição do NPS no estádio com novos line-arrays. De [8].

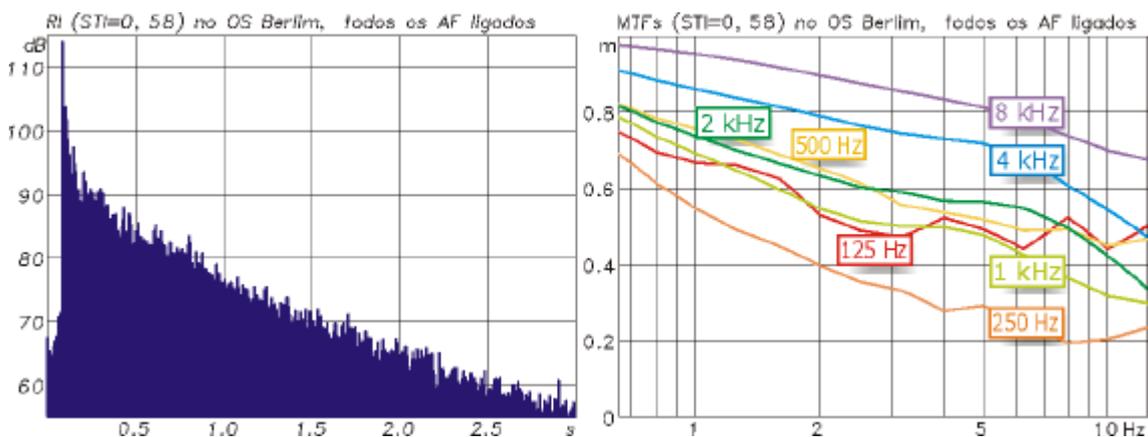


Fig. 16: Resposta impulsiva e MTFs no estádio Olympia, Berlim.

### 10.3 Hangar de dirigíveis perto de Berlim

O terceiro exemplo foge um pouco do escopo de lugares comuns a serem sonorizados. Trata-se de um hangar gigante que foi erguido a 80 km de Berlim para abrigar a construção de dirigíveis para transporte de cargas. A empresa faluiu e o hangar agora abriga um parque temático tropical (bom, pelo menos não virou templo evangélico).

As medições foram feitas na fase da transição do destino da edificação enorme, com comprimento de 360 m, largura de 240 m e uma cúpula cujo pé-direito atinge 107 m no centro.



Fig. 17: Hangar para a montagem de dirigíveis com palco para um evento.

A RI mostra uma cornucópia de reflexões chegando centenas de milissegundos atrás do som direto, com intensidade bem maior do que ele. A forma côncava tanto das paredes quanto do teto surte um fortíssimo efeito focal que amplifica os ecos.

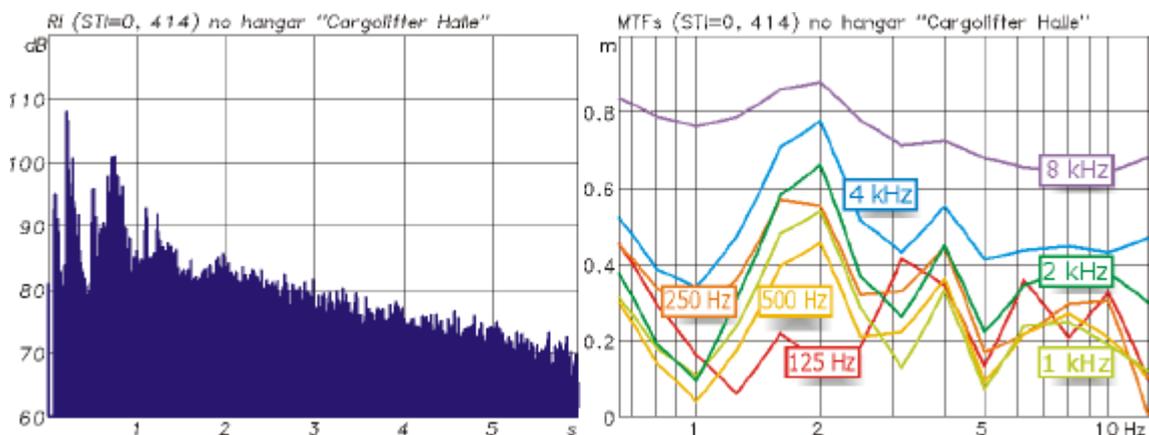


Fig. 18: Resposta impulsiva e MTFs no hangar "Cargo Lifter".

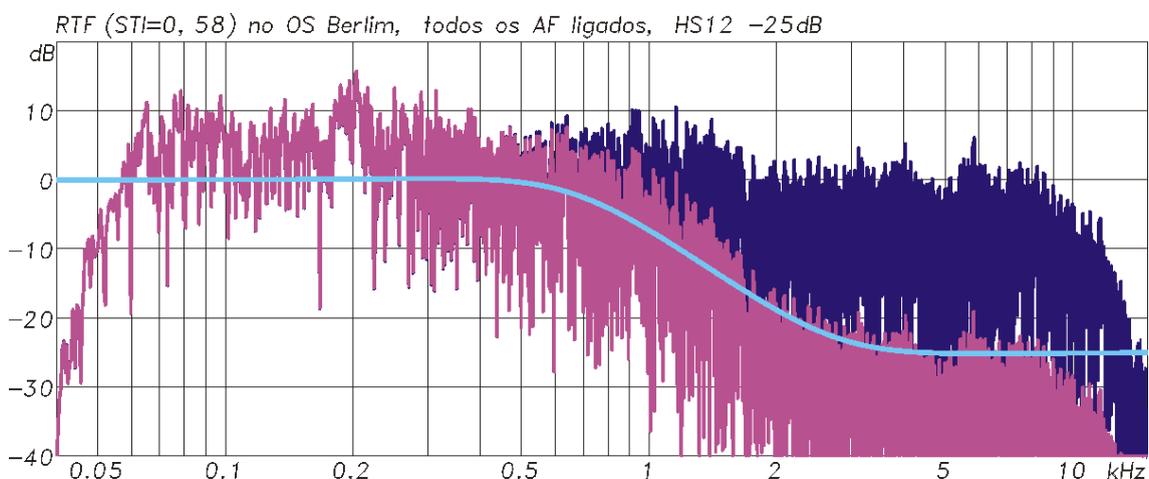
O STI calculado a partir dessa RI descomunal é de 0.414, praticamente o mesmo valor também calculado para o centro cultural vazio no primeiro exemplo. Isso podia nos levar a crer que embora tendo uma avaliação “fraco”, ainda poderia ser possível entender algo nesse recinto. No entanto, testes de auralização com base nesta RI mostram que qualquer segmento de fala volta completamente embaralhado, com inteligibilidade praticamente nula. Esse é um bom exemplo para o fato que o STI trata reflexões tardias e fortes de forma demasiadamente generosa, e não leva suficientemente em conta a relação do nível do som direto com elas.

## 11. Fraquezas do STI

Embora normalmente exiba uma correlação boa com a inteligibilidade percebida, o STI pode errar bastante em casos extremos. O último exemplo é testemunha disso. De forma geral, o STI não se importa de forma adequada com a estrutura temporal das respostas impulsivas. Por exemplo, um resultado idêntico para o valor de STI é obtido entre uma RI e a sua versão invertida no eixo temporal, ação que produz uma inteligibilidade desastrosa! Aparentemente, algumas propriedades do nosso senso auditivo, que aproveita componentes chegando no máximo 50 ms após o som direto, mas é despistado bastante por componentes chegando mais tarde, não são modeladas de forma psicoacusticamente correta no cálculo do STI.

Relacionado a esse aspecto é a percepção tridimensional do nosso ouvido, que ajuda bastante para discernir uma voz numa barafunda de outras vozes ou ruído em geral. Porém, o STI normalmente é levantado com a ajuda de um microfone onidirecional, desconsiderando as direções das quais os componentes da RI chegam às orelhas do ouvinte e as diferenças interaurais. Por isso, seria mais adequado adquirir RIs binaurais com uma cabeça artificial [6]. Mas esse é um problema que também vale para a avaliação dos demais parâmetros acústicos.

Outro ponto fraco do STI diz respeito ao equilíbrio tonal da transmissão. Em ambientes onde a relação S/R não é problema, o STI pouco se importa como a resposta em frequência de um sistema de sonorização. Um exemplo típico e com bastante relevância prática é a falta de agudos. No exemplo da Fig. 19, a função de transferência medida no estádio em Berlim foi deliberadamente reduzida em até 25 dB a partir de 1 kHz.



**Fig. 19: Função de transferência no estádio Olympia em Berlim, simulação com e sem perda aguda de agudos (curva azul clara: o filtro empregado).**

Como o nível não tem influência nos componentes do STI que dizem respeito à acústica de sala (reflexões e reverberação), os fatores **m** da matriz praticamente não mudaram. Portanto, o valor final do STI (segundo a definição original) também não sofre alterações. A correção do mascaramento introduzida na terceira edição da norma só traz uma pequena queda do STI de 0.58 para 0.55. De maneira alguma acusa de forma adequada a perda quase total de inteligibilidade neste exemplo. E não se trata de um caso puramente teórico. Ele pode acontecer de forma semelhante quando os tweeters do sistema de sonorização forem queimados. Como se vê neste exemplo, essa ocorrência praticamente não poderia ser detectada por uma medição do STI.

O STI não é somente destinado aos efeitos naturais de acústica de sala e ruído de fundo. Já pelas origens, quando os idealizadores visaram quantificar com meios objetivos a qualidade de transmissão de rádios de comunicação, ele sempre teve a pretensão de também considerar corretamente qualquer influência técnica na transmissão de som, seja ela deliberada ou não. Porém, nem todos os efeitos de processamento e sonorização de fala têm reflexo correto sobre o STI. Alguns até mudam o STI na direção errada.

O exemplo mais notável disso é o uso de compressores. Um compressor aumenta o nível nos trechos baixos de fala e diminui os picos. Isso tem duas conseqüências para o cálculo do STI. A primeira é que o método indireto, chegando ao STI através da RI, não pode ser aplicado. As envoltórias dos sinais de excitação têm uma característica completamente diferente em comparação à fala, surtindo reações do compressor igualmente diferentes. No caso de varreduras de seno, o compressor não mudaria o ganho, pois o nível já se encontra em patamar constante (o negócio é diferente para compressores de multi-banda, é claro). Portanto, a inclusão do compressor não surtiria efeito nenhum. No caso de ruído pseudo-aleatório (como MLS), muito sensível a quaisquer mudanças não-lineares e/ou variáveis no tempo, o uso do compressor iria destroçar a RI, aumentando dramaticamente o ruído de fundo aparente.

Então, transmissões de fala que passam por um compressor só podem ser investigadas com o método direto, quer dizer tocando seqüencialmente os 98 ruídos filtrados e modulados. Porém, como o compressor diminui os picos e levanta os vales, o que corresponde a uma diminuição da profundidade de modulação, o STI sofrerá uma queda notável, acusando uma perda de inteligibilidade. No entanto, o contrário é verdade: o compressor (empregado na medida certa) melhora a inteligibilidade e em casos críticos com muito ruído de fundo, pode elevar uma transmissão do ininteligível até o razoável.

Uma coisa parecida acontece se *frequency-shifters* são aplicados para suprimir realimentações. O deslocamento linear de frequência em torno de 5 ou 10 Hz não prejudica a inteligibilidade, mas tem efeito bastante negativo no STI, pois as frequências centrais de modulação não são mais acertadas.

Existe uma série de diversos erros que podem acontecer em equipamentos eletro-acústicos que não têm efeito adequado no STI.

Um exemplo é o chamado *center-clipping* (ceifa central). Ele significa que o sinal, ao mudar a polaridade, desaparece completamente abaixo de um certo limite. Esse efeito pode ser

ocasionado nos estágios complementares de amplificadores de potência se a *corrente de repouso* for baixa. Também pode se manifestar em conectores com mau contato e circuitos eletrônicos com defeito em geral. Como a intensidade do sinal já é baixa nos trechos que desaparecem, esse fenômeno muito desagradável tem pouquíssima influência no cálculo do STI. Todavia, prejudica bastante a inteligibilidade.

Outro exemplo é transmissão intermitente, que pode acontecer em sistemas sem fio quando o sinal RF cair abaixo do limiar do *squelch* (supressor de ruído quando o canal está vazio), ou em sistemas digitais quando houver qualquer erro de transmissão. *Drop-outs* na reprodução de fitas velhas, arquivos danificados, ou de *streams* de áudio através de uma conexão digital compartilhada e temporariamente sobrecarregada são outros exemplos corriqueiros dessa condição desagradável. No entanto, esse tipo de estorvo também não incomoda o STI. Uma simulação típica é de periodicamente ligar e desligar o som, com silêncio de 100 ms em cada período. Esse estopício de fala a deixa quase completamente ininteligível. Por incrível que pareça, o STI aumenta ligeiramente nessa situação!

Como se viu nos capítulos 4 a 6, o cálculo do STI é bastante complexo. Tendo em vista todas essas situações nas quais o STI pode falhar, é justificável a dúvida se todo o esforço realmente valha a pena. Bradley [7] até mostrou que uma medida bem mais simples, o  $U_{50}$  (que é o *useful-to-detrimental ratio*, ou seja, a *claridade*  $C_{50}$  com influência do ruído de fundo), tem uma correlação tão boa com a inteligibilidade quanto o STI.

## 12. Conclusões

O STI certamente não é uma panacéia na análise e predição da inteligibilidade em lugares padecendo ruído de fundo e acústica de sala carente, mas ele é bem prático e certamente um bom primeiro indicador de problemas que comprometem o entendimento. Embora possa falhar fatalmente em situações extremas e um tanto atípicas, ele geralmente tem uma correlação satisfatória com a inteligibilidade de fato experimentada no lugar pesquisado.

Uma pré-condição para lugares que dependem da amplificação da voz é que o sistema de sonorização funcione sem erros técnicos, sendo bem equalizado e livre de certos tipos de processamento (notavelmente dispositivos que mudam a dinâmica do sinal, tais como limitadores, compressores e *gates*). A resposta em frequência sempre tem que ser avaliada separadamente (mas podendo utilizar convenientemente as RIs já medidas para avaliar STI), porque o STI é insensível a desequilíbrio tonal.

O lado positivo do STI é que ele é uma ferramenta padronizada e amplamente utilizada mundialmente. Em casos litigiosos, ele pode desempatar um laudo técnico. Ao mesmo tempo, isso significa que ele tem que ser levantado cautelosamente, sempre visando excluir possíveis fontes de erros. A maior dela sem dúvida é a estipulação da relação S/R. O ruído depende muito da situação e o nível do sinal também pode variar bastante entre locutores diferentes. Em contrapartida, o componente da acústica de sala pode ser traçado com boa precisão, sobretudo se o sistema de sonorização a ser utilizado na reprodução da voz do locutor também for utilizado para medir as respostas impulsivas.

A avaliação automatizada do STI por um equipamento não substitui um especialista em acústica, muito pelo contrário. Com um pouco de experiência, a matriz dos valores  $m$  (ou melhor, as MTFs) informam qual é a fonte predominante que compromete a inteligibilidade. Cabe então a engenheiro tarimbado identificar a origem do problema e tirar ele pela raiz, se for factível.

### 13. Referências

- [1] IEC 60268-16:2003, “Objective rating of speech intelligibility by speech transmission index”.
- [2] Past, present and future of the Speech Transmission Index, TNO Human Factors, Soesterberg, Netherlands (ISBN 90-76702-02-0).
- [3] TAMMO HOUTGAST, HERMAN J.M. STEENEKEN (1971), “Evaluation of Speech Transmission Channels by Using Artificial Signals”, *Acústica* 25, 355-367.
- [4] HERMAN J.M. STEENEKEN, TAMMO HOUTGAST, “Basics of the STI measuring method”, available at <http://www.steeneken.nl/sti.html>
- [5] HERMAN J.M. STEENEKEN, TAMMO HOUTGAST, “Improvements of STI: frequency weighing, gender, level dependent masking, and phoneme specific prediction” available at <http://www.steeneken.nl/sti.html>
- [6] PETER MAPP, “Systematic & Common Errors in Sound System STI and Intelligibility Measurements”, 117<sup>th</sup> AES 2004 San Francisco, Preprint 6271.
- [7] JOHN S. BRADLEY,” Optimising Sound Quality for Classrooms”, XX Encontro da SOBRAC, Rio de Janeiro, Outubro de 2002
- [8] ANSELM GOERTZ, CHRISTIANE BANGERT, WOLFGANG AHNERT, STEFAN FEISTEL,” Setup and final measurements for a PA-System at the example of a large sports stadium in Berlin”, Tonmeistertagung Leipzig, Novembro de 2004, (infelizmente, só em alemão)
- [9] G. LEMBRUGGEN, A. STACEY, “Should the Matrix be reloaded?” Proceedings of the Institute of Acoustics, Vol. 25. Pt.8, 2003