

RUMSAKUSTISK KVALITET BEDÖMD AV TALARE

Alf Berntson
Chalmers tekniska högskola
Avd för teknisk akustik
412 96 GÖTEBORG

1. Sammanfattning

I två olika subjektiva prov bedömdes talbekvämligheten i simulerade ljudfalt. Ljudfalten i den första undersökningen simulerade tre olika tvarsektioner med mycket olika bredd-höjd-forhållanden. I den andra undersökningen simulerades olika avstånd till väggen bakom talaren. Alla reflexer med en fordroningstid mindre än 125 ms simulerades i de totalt nio olika ljudfalten. Med hjälp av flerdimensionell skalning av de parvisa preferensbedomningarna erhöles två underliggande dimensioner. Denna tvådimensionella psykometriska modell förklarade över 90 % av den totala variansen i parjämförelsebedomningarna. En grundlig teoretisk analys av ljudfalten med hjälp av spegelkällometoden indikerade att de flesta försökspersonerna föredrog hög total nivå (avstånd till väggar) samt låg "temporal diffusitet" (tidsmässig spridning och jämnhet) för impulsresponsen. Observera att talarna föredrog färgad klang i "rums-svaret" troligen på grund av att direktljudets (mun till öra) maskering av reflexerna i dessa fall blev lagre, vilket gav bättre hörbarhet av det reflekterade ljudet.

2. Experiment 1

I detta prov ville vi undersöka hur bredd-höjd-forhållandet i en sal påverkar talaren. För musiklyssning har flera undersökningar visat att tidiga laterala reflexer är viktiga för hög subjektiv kvalitet medan däremot tidiga (starka) vertikala reflexer kan ge en onjuter klangfärgsforändring. Mot denna bakgrund är det intressant att undersöka vikten av bredd-höjd-forhållandet för musiker, sångare och talare.

Tre olika tvärsektioner med en symmetrisk (mitt emellan sidoväggarna) och en asymmetrisk (förhållandet 3/5 mellan avstånden till sidoväggarna) talarposition i var och en av sektionerna simulerades med hjälp av en elektroakustisk simuleringsanläggning (se figur 1 och 2). Väggen framför och väggen bakom talaren var helt absorberande, övriga fyra ytor totalt reflekterande. De simulerade "rummen" kan alltså betraktas som oändligt långa rektangulära korridorer. Reflexer med en fördröjningstid mindre än 125 ms, motsvarande ca 50 spegelkällor, simulerades med hjälp av 9 st högtalare i halvekofritt rum. Ljudfälten beräknades med spegelkällemetoden och data för inställning av simulatorn lagrades i dator. På grund av de förhållandevis små hörbara skillnaderna användes parjämförelsemetoden med fyra replikationer av varje par. Tio personer bedömde alltså vilket fall (A eller B) i varje par de tyckte var "bekvämast att tala i". Omkopplingen mellan stimuli gjordes på mindre än en sekund (inställning av 1000 datorstyrda dämpsatser och 13 datorstyrda fördröjningsenheter) och kontrollerades helt och hållet av försökspersonerna. Även insamlingen av svaren sköttes automatiskt av datorn. En grundligare beskrivning av testproceduren och utrustningen finns i referens 1 och 2.

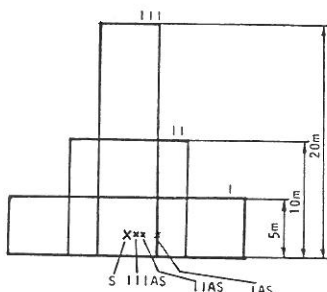


Fig.1 De tre tvärsektionerna och de symmetriska (S) och asymmetriska (AS) talarpositionerna.

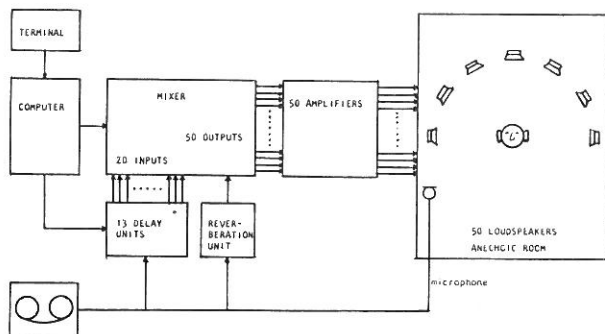


Fig.2 Block-schema över simulatorn. Efterklangsenheten användes ej i experimenten.

3. Experiment 2

En naturlig fortsättning efter experiment 1 var att undersöka hur avståndet till väggen bakom talaren påverkar talbekvämligheten. En av tvärsektionerna i experiment 1 användes (fall IIAS) och 5 olika avstånd till bakomvarande väggen enligt figur 3 undersöktes. I dessa ljudfält uppkommer ytterligare ett plan av spegelkällor jämfört med ljudfälten i experiment 1, dvs ca 100 reflexer upp till 125 ms fördröjning (se figur 4). I detta prov användes därför 29 av de totalt 50 högtalarna i simuleringsanläggningen. Testproceduren för de 15 försökspersonerna var densamma som i experiment 1.

4. Resultat

Parjämförelserna analyserades parvis varvid signifikanta skillnader togs fram (ref 1 och 2). Endimensionell skalning enligt "Thurstone's Case V model" (ref 1, 2 och 3) samt flerdimensionell skalning (ref 4) utfördes också. Flerdimensionell skalning av preferensbedömningar kan göras med hjälp av datorprogrammet MDPREF (Multi Dimensional PREFERENCE scaling). MDPREF ger en n-dimensionell rymd där både stimuli och försökspersoner finns representerade. Försökspersonerna representeras av vektorer och stimuli av punkter. Projektion av stimuli på vektorerna ger de individuella relativa preferensskalorna.

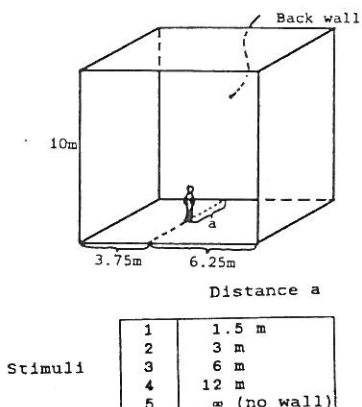


Fig.3 Det simulerade rummet i experiment 2. Väggen framför talaren är helt absorberande.

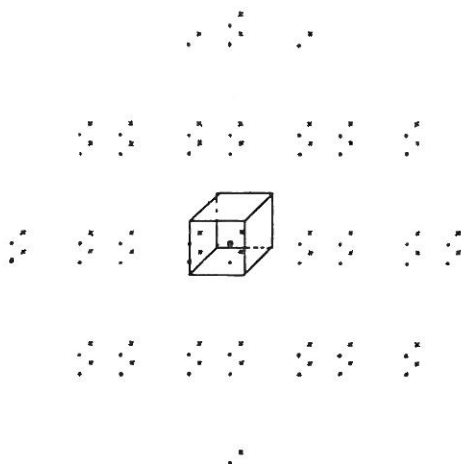


Fig.4 I experiment 2 fås två parallella plan av spegelkällor på avståndet 2a från varandra (prickar och kryss).

Figur 5 och 6 visar de tvådimensionella skalningarna för de två experimenten. Efter inspektion av de fyra replikationerna uteslöts 2 av de 10 försökspersonerna i experiment 1 och 3 av de 15 försökspersonerna i experiment 2 på grund av låg reliabilitet och/eller dålig diskriminationsförmåga. Tas dessa personer med fås sämre anpassning av modellen till indata och skalningen blir mer opålitlig.

Skalningen av bedömningarna i experiment 1 visar en tydlig preferens för de symmetriska talarpositionerna. Den tvådimensionella skalningen förklarar 91.7 % av variansen i data vilket är fullt tillräckligt. Högre dimensioner avspeglar endast den slumpmässiga spridningen i indata. I figur 5 ser vi att dimension 1 står för den största delen av variansen (74.1 %) och bör kunna tolkas i termer som "klangfärgsförändring", "fladdereko", "avvikelse från exponentiellt efterklangsförlopp", etc. (Se nästa avsnitt). Föredrar man tidiga laterala framför tidiga vertikala reflexer? Ur parjämförelsematrisen erhöles endast en signifikant (5 %) skillnad mellan tvärsektionerna; den smala bättre än den breda i de asymmetriska talarpositionerna (IIIAS > IAS).

Skalningen av data i experiment 2 visar att dimension 1 står för en mycket stor del av variansen (82.3 % av 95.4 %). Utefter dimension 1 skiljs framför allt stimulus 1 från stimulus 5. Den största skillnaden mellan dessa stimuli är total reflekterad nivå. De flesta föredrar alltså starka reflexer, men vi ser också att åtminstone tre personer har en annan åsikt. Dessa individuella skillnader beror troligen mest på skillnader i talnivå. Har man en stark röst ger stimulus 1 för kraftig respons som upplevs negativt av dessa personer.

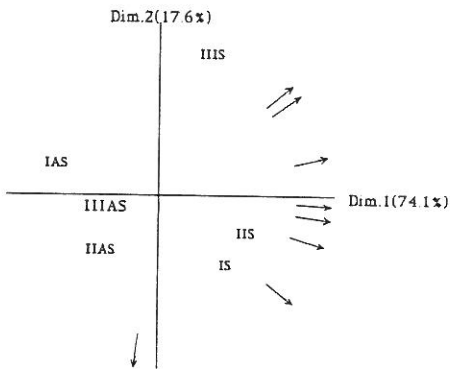


Fig. 5 Experiment 1. Tvådimensionell MDPREF-skalning.

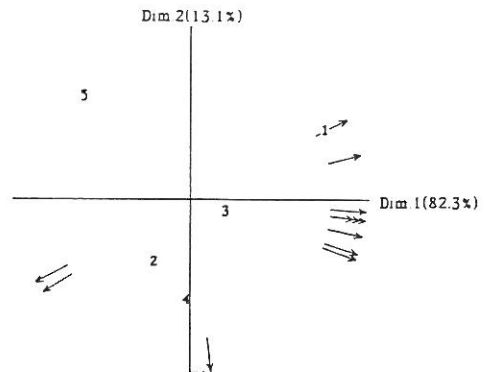


Fig. 6 Experiment 2. Tvådimensionell MDPREF-skalning.

5. Teoretisk analys av ljudfälten

5.1. Inledning

En talare påverkas kraftigt av medhörningen. Ur ljudproducerande synvinkel är vi människor ett återkopplat reglersystem, men inte enbart akustiskt utan även fysiologiskt-mekaniskt (att det "kanns" bra). Exempelvis påverkas vi starkt av medhörningen när vi talar i telefon (kallas sidoton på "telefonispråk"). Hör vi oss själva dåligt höjer vi rösten och vice versa. I ett rum blir "medhörning" från det reflekterade ljudet fördrojt och betydligt svagare än direktljudet från mun till öra (utanför och inuti huvudet). En orsak till att vi trots allt uppfattar stora skillnader mellan olika rum är att vi inte producerar en stationär signal. På grund av modulationen i talet maskeras reflexer som är fördrojda betydligt mindre än tidiga reflexer speciellt om nivån är i samma storleksordning.

5.2. Spegelkällemetoden

Nivå- och tidsfördelningen för reflexerna i de simulerade rummen kan beräknas med spegelkällemetoden. Ersatt talaren med en punktkälla. Ljudtryckets avståndsberoende från en punktkälla beskrivs av Greens funktion i fritt fält, denna ger att

$$p \sim \frac{e^{-jkr}}{r}$$

som alltså beskriver fas och amplitudförändring som funktion av enbart avståndet. Genom spegling i alla ytor beräknas alla spegelkällors läge upp till en viss fördröjningstid.

Ljudtrycket i en punkt i rummet relativt ljudtrycket i fritt fält 1 m från punktkällan fås då genom summering av de n st bidragen

$$p(\omega, r_k, r_m) = \sum_{i=1}^n R_i(\omega) \frac{e^{-jkr_i}}{r_i}$$

dar r_k = källans positionsvektor

r_m = mottagarpunktens positionsvektor

r_i = avståndet mellan spegelkälla i och mottagarpositionen

$R_i(\omega)$ = produkten av reflektionsfaktorerna för spegelkälla i

$$k = \frac{\omega}{c}$$

Efter modellen ovan skrevs ett datorprogram som beräknar det komplexa ljudtrycket i ett rektangulärt rum. Reflektionsfaktorerna för de maximalt sex ytorerna är komplexa men i nuvarande preliminära version är de frekvensoberoende.

5.3. Överföringsfunktioner - impulsresponser

Överföringsfunktionen i rummet beräknades i spegelkälleprogrammet för ett antal frekvenser. I beräkningarna som presenteras här har 2048 frekvenslinjer använts och upplösningen var $\Delta f = 8$ Hz. Detta ger $N = 4096$ sampel i tidsfunktionen (invers Fouriertransform av det beräknade komplexa spektrumet, alltså rummets impulsrespons) med en fönsterlängd $T = \frac{1}{\Delta f} = 125$ ms. Bandbredden blev alltså $B = \Delta f \cdot \frac{N}{2} = 16\,384$ Hz, eller annorlunda uttryckt, den obegränsade överföringsfunktionen i rummet multiplicerades med ett idealt lågpasfilter med gränshfrekvensen 16 384 Hz. Denna lågpasfiltrering ger naturligtvis upphov till utbredda pulser i impulsresponsen. Ett lågpasfilter med överföringsfunktionen

$$H(f) = \begin{cases} a, & -B \leq f \leq B \\ 0 & \text{annars} \end{cases}$$

ger impulssvaret

$$h(\tau) = 2 a B \left(\frac{\sin 2\pi B \tau}{2\pi B \tau} \right)$$

I det diskreta fallet är ju impulsresponsen samplad, om då fördröjningstiden för en reflex hamnar exakt på ett sampel (τ jämnt delbart med samplingsintervallet T/N) kommer endast en linje att synas för denna reflex (jämför figur 7). Om vi däremot hamnar med τ någonstans mellan två sampel kommer vi att få något som liknar en $\frac{\sin x}{x}$ -funktion, amplituden för reflexen (med enkel linjär interpolation mellan samplen) kommer då att underskattas. För att minska

detta fel sattes alla frekvenskomponenter över 6 kHz till noll (men med samma upplösning Δf och lika många element som tidigare) innan inverstransformeringen. Detta resulterade i en breddning av pulserna och därmed minskade amplitudfelet (se figur 8). Ett enklare tillvägagångssätt är att utgå från de beräknade fördröjningstiderna från spegelkällaberäkningen och skifta impulserna så att de hamnar exakt på ett sampel. Men om antalet sampel är begränsat och impulssvaret långt blir resultatet att man inför betydande fasfel. (Att minimera fasfelen är i vissa fall mycket viktigt t ex vid binaurala simuleringar från beräknade impulssvar (ref [5])). Dessutom om man har komplexa reflektionsfaktorer (vinkelberoende komplicerar) är beräkning i frekvensplanet enklare, därför överväger fördelarna för detta beräkningssätt.

I figur 9 visas tre av de beräknade impulsresponserna. Eftersom källa och mottagare var i samma punkt är endast reflexerna medtagna.

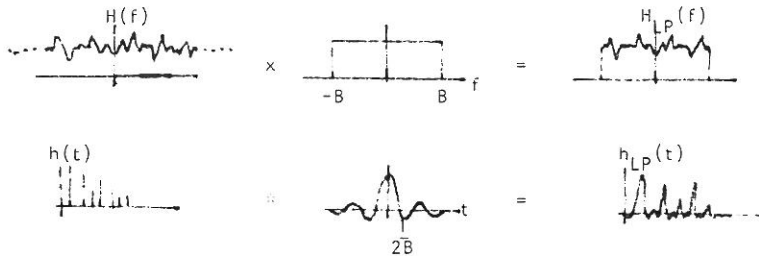


Fig.7 Genom LP-filtrering av överföringsfunktionen breddas pulserna i impulsresponserna. (jämför fig.8)

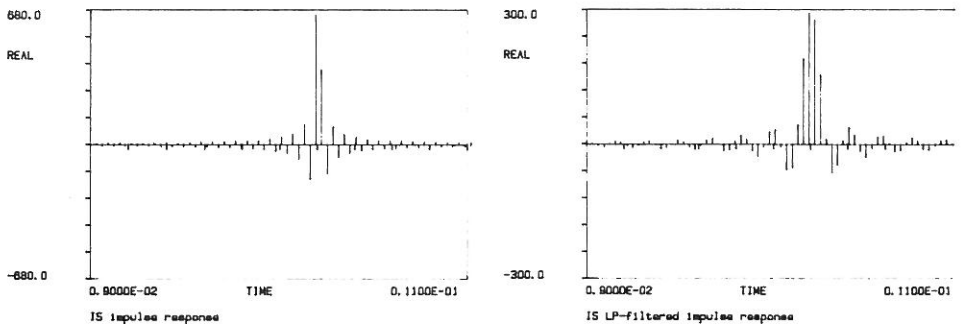


Fig.8 Fler sampel per puls ger mindre amplitudfel vid linjär interpolation mellan sampel.

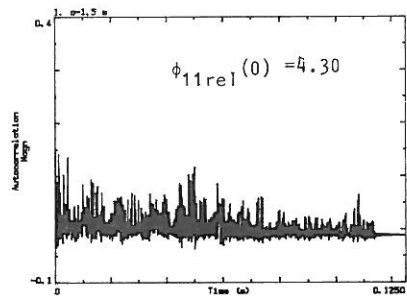
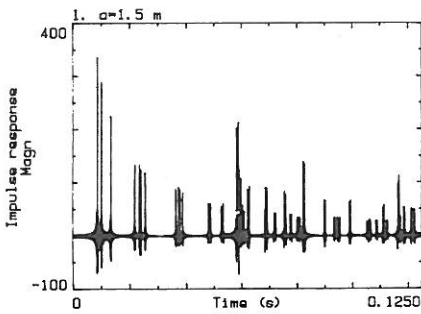
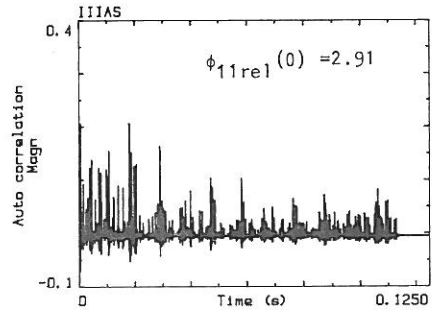
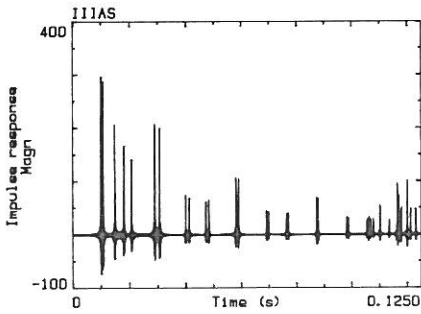
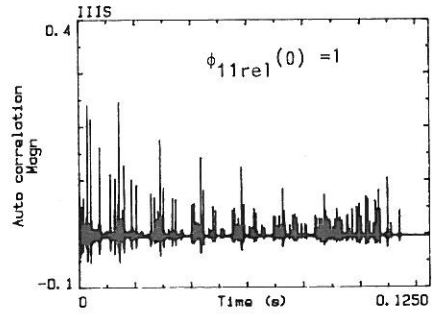
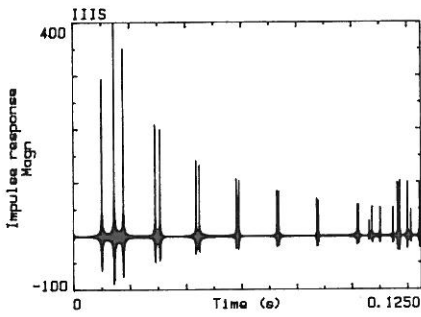


Fig.9 Lp-filtrerade (6kHz) impulsresponser och autokorrelationer.

5.4. Autokorrelation och korskorrelation i rum

5.4.1. Teori

Låt $s_1(t)$ och $s_2(t)$ vara två stationära tidsfunktioner, då är

$$\phi_{12}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T s_1(t) \cdot s_2(t + \tau) dt \quad (1)$$

korskorrelationen för $s_1(t)$ och $s_2(t)$.

Om $s_1 = s_2$ fås autokorrelationen

$$\phi_{11}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T s_1(t) \cdot s_1(t + \tau) dt$$

Låt nu $g(t)$ vara impulsresponsen för transmissionen mellan källan och en punkt i ett rum. Om vi vill utföra korrelationsberäkningar i rummet med hjälp av beräknade impulsresponser $g(t)$ kan vi inte direkt stoppa in dessa i uttrycken ovan eftersom $g(t)$ inte är stationära funktioner. Antag istället att vi exciterar rummet med en stationär stokastisk brusignal $r(t)$ med stor bandbredd, då blir responsen i två punkter i rummet

$$s_1(t) = \int_0^\infty g_1(x) r(t - x) dx \quad (\text{fäلتning})$$

$$s_2(t) = \int_0^\infty g_2(y) r(t - y) dy$$

Dessa funktioner är stationära och kan därmed sättas in i ekv (1)

$$\begin{aligned} \phi_{12}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left[\int_0^\infty g_1(x) r(t-x) dx \right] \left[\int_0^\infty g_2(y) r(t + \tau - y) dy \right] dt \\ &= \int_0^\infty g_1(x) dx \int_0^\infty g_2(y) dy \cdot \left(\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T r(t - x) \cdot r(t - y + \tau) dt \right) \end{aligned}$$

Eftersom $r(t)$ är stationär stokastisk brussignal med stor bandbredd kommer uttrycket innanför parentesen att ge en deltafunktion då $t - x = t - y + \tau \Leftrightarrow y = x + \tau$, alla övriga argument ger försumbara bidrag.

Då fås

$$\phi_{12}(\tau) = \int_0^{\infty} g_1(x) \cdot g_2(x + \tau) dx$$

Vi kan alltså använda impulsresponserna direkt för att beräkna auto- och korskorrelationerna. Observera att $\lim_{T \rightarrow \infty} \frac{1}{T}$ ej finns med framför integralen i denna ekvation.

5.4.2 Tillämpning av autokorrelation

Med hjälp av autokorrelationsfunktionen kan man få ett mått på den "temporala diffusiteten" (ref 6), dvs den tidsmässiga spridningen av reflexerna i impulsresponserna. Periodiskt upprepade reflexer kan ge klangfärgsförändring för korta periodtider ($\Delta\tau$ ungefär < 25 ms) eller fladdereko för längre periodtider. I impulsresponser kan periodiska reflexer döljas i andra slumpmässigt fördelade reflexer även om den subjektiva effekten är fullt hörbar. Här har autokorrelogrammet den fördelen att de "slumpmässigt" fördelade reflexerna mest bidrar till värdet $\phi_{11}(0)$ och de periodiska visar sig som toppar i $\phi_{11}(\Delta\tau)$. Med samplad impulsrespons (N st sampel) fås

$$\phi_{11}(0) = \sum_{i=1}^N h_i^2$$

där

h_i = impulsresponsens värde för sampel i

$\phi_{11}(0)$ är alltså ett mått på den "totala kvadrerade impulsamplituden". I figur 9 visas de normaliserade autokorrelationsfunktionerna $\phi_{11}(\tau)/\phi_{11}(0)$, dvs värdet för $\tau = 0$ är alltid 1. $\phi_{11}(0)$ för de olika fallen anges relativt fall IIIS.

Enstaka starka reflexer kan också ge upphov till "färgat" ljud eller hörbart eko, även detta visar sig som en topp i autokorrelogrammet.

5.4.3. Tillämpning av korskorrelation

I analogi med autokorrelationen ger korskorrelationen ett mått på den "rumsliga diffusiteten". Korskorrelationen mellan öronen har visat sig ha samband med subjektiva sensationer som har fått många namn : "spatial impression, spatial responsiveness, ambience, apparent source width, feeling of envelopment, auditory spaciousness" (ref 7).

IACC (Inter Aural Cross-Correlation (ref 8)) är det objektiva mått som har använts mest för att uppskatta dessa subjektiva fenomen. IACC definieras som maximum för beloppet av den normaliserade korskorrelationen mellan öronen för $|\tau| < 1$ ms,

$$IACC = |\phi_{12n}(\tau)|_{\max}, \quad |\tau| < 1 \text{ ms}$$

där

$$\phi_{12n}(\tau) = \frac{\phi_{12}(\tau)}{\sqrt{\phi_{11}(0) \cdot \phi_{22}(0)}}$$

Ju lägre värden på IACC desto högre subjektiv "rymdkänsla" etc.

I figur 10 visas två exempel på beräknade normaliserade korskorrelationsfunktioner. Som tidigare är beräkningarna gjorda från de låg-passfilterade (6 kHz) överföringsfunktionerna. Källpositionerna var desamma som tidigare medan mottagarpositionerna var 8 cm på var sin sida om källpositionerna, motsvarande mun och öron (naturligtvis fattas huvudet i beräkningarna). Man ser tydligt i figuren att den symmetriska källpositionen ger högre värden. I den asymmetriska positionen framträder en topp vid $\tau \approx \pm 0.46$ ms som motsvarar tidsskillnaden mellan mottagarpositionerna för rent laterala reflexer. Detta verkar rimligt eftersom de laterala reflexerna är starka och ej infaller samtidigt som i det symmetriska fallet. Observera att talaren föredrog de symmetriska talarpositionerna som alltså gav högre IACC-värden. Detta beror troligen på direktljudets kraftiga maskering av det mer "diffusa" ofärgade ljudet i de asymmetriska positionerna.

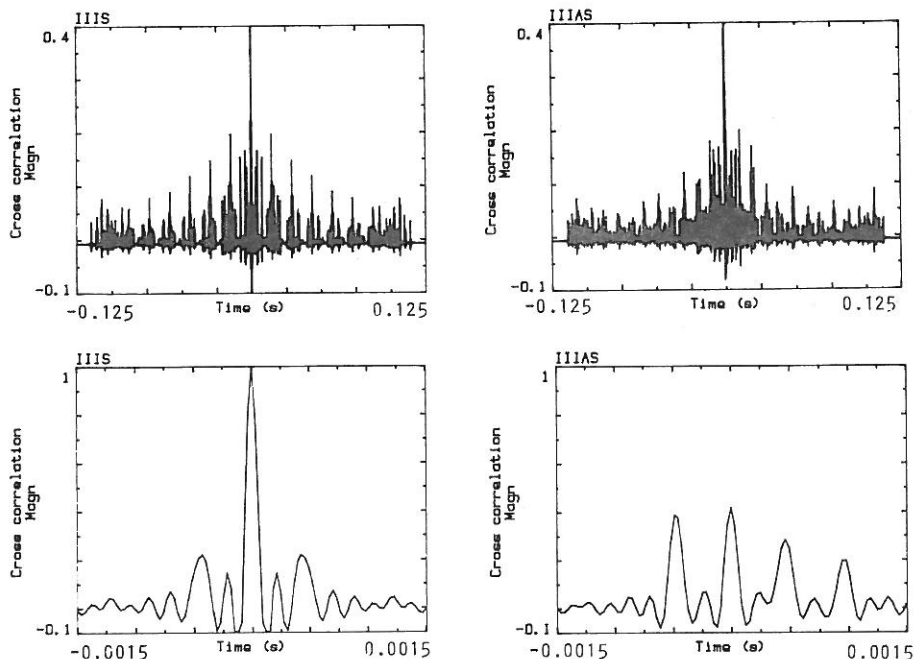


Fig.10 Normaliserade korskorrelationsfunktioner.

5.5. Överföringsfunktioner

Om man jämför de beräknade överföringsfunktionerna (smalbandiga $\Delta f = 8$ Hz) är det mycket svårt att se någon systematisk skillnad förutom nivån. Tersbandsfiltrering kommer att utföras och kan eventuellt indikera systematiska skillnader, men skall man efterlikna örat bör man snarare beräkna korttidspektrum (short-time spectrum (ref 9)), dvs införa någon slags vägning av impulsresponserna. Atal fann en vägningsfunktion enligt figur 11 grundad på tröskelvärden för subjektiv färgning av kamfilterfiltrerade stationära brussignaler. En talare, musiker eller sångare producerar inte en stationär signal och direkt ljudet från mun till öra kommer att maskera de tidiga reflexerna kraftigt. Det bör därför vara rimligt att modifiera vägningsfunktionen så att den allra tidigaste delen vägs mycket lite. Hur denna vägningsfunktion bör se ut kräver dock fortsatta undersökningar.

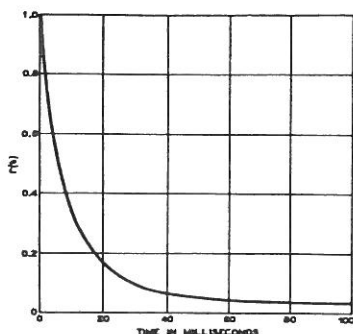


Fig.11 Vagningsfunktion för
beräkning av korttids-
spektrum enligt Atal.

Referenser

- [1] M. Kleiner, A. Berntson: Early frontal plane reflections preferred for talkers. Proc. 12 ICA (1986).
- [2] A. Berntson: Preferred distance to a wall behind talkers. Proc. NAM (1986).
- [3] W.S. Torgerson: Theory and methods of scaling. John Wiley, New-York (1958).
- [4] A. Berntson: Talking comfort in auditoria. Interpretation of subjective judgements using multidimensional scaling. Proc. 6. FASE (1986).
- [5] P.M. Peterson: Simulating the response of multiple microphones to a single acoustic source in a reverberant room. JASA 80(5) (1986).
- [6] H. Kuttruff: Über Autokorrelationsmessungen in der Raumakustik. Acustica vol 16 (1965/66).
- [7] J. Blauert, W. Lindemann: Some further psychoacoustic analysis. JASA 80(2) (1986).
- [8] Y. Ando: Calculation of subjective preference at each seat in a concert hall. JASA 74(3) (1983).
- [9] B.S. Atal, M.R. Schroder, K.H. Kuttruff: Perception of coloration in filtered gaussian noise. Short-time spectral analysis by the ear. Proc. 4 ICA (1962).