

Spraakverstaanbaarheid

Kennisbank Bouwfysica

Dictaat ct4220 Bouwfysica II, TU-Delft Faculteit Civiele Techniek en Geowetenschappen,
samengesteld door prof.ir. J.J.M. Cauberg

1 Inleiding

Het is voor iemand die in een zaal naar een spreker luistert van primair belang dat het gesprokene duidelijk en gemakkelijk kan worden verstaan. Daarnaast is het wenselijk dat de luisteraar het geluid ervaart als komend van de plaats waar de spreker zich bevindt.

De spraakverstaanbaarheid hangt af van de spreker (stemvolume, articulatie), de luisteraar (oorgevoeligheid) en de ruimteakoestische parameters zoals die worden gespecificeerd door de pulsresponsie tussen spreker- en luisteraarpositie.

Spraakverstaanbaarheid is in principe gedefinieerd als het percentage gesproken woorden die door de luisteraar correct wordt verstaan. Als “woorden” worden bij spraakverstaanbaarheidsproeven meestal zgn. logatomen gebruikt. Dit zijn éénlettergrepige combinaties van een klinker met medeklinkers ervoor en erna, die niet per se een bestaand woord hoeven te vormen, (bij voorbeeld duk, zwat, keeg). Dikwijls wordt bij het samenstellen van logatomenlijsten rekening gehouden met de frequentie waarin de klanken in het normale spraakgebruik voorkomen.

Tijdens proeven worden de logatomen uitgesproken, ingebed in “draagzinnen”, bij voorbeeld: eerst hoort u “duk”, dan zeg ik “zwat”, vervolgens “keeg”, etc.. Wanneer de resultaten van dit soort proeven worden gemiddeld over een aantal “normale” sprekers en luisteraars, volgt “de” spraakverstaanbaarheid voor een bepaalde spreker- en luisteraarpositie.

Door de proeven te herhalen voor verschillende posities kan de spraakverstaanbaarheid in een ruimte in kaart worden gebracht.

2 Speech Transmission Index (STI)

Bovenstaande procedure is erg omslachtig en tijdrovend. Daarom hebben vele onderzoekers, ook in Nederland, getracht fysische parameters te definiëren die een maat zijn voor de spraakverstaanbaarheid. Houtgast en Steeneken introduceerden de “Speech Transmission Index” (STI), gebaseerd op de analogie tussen lopende spraak en intensiteitsgemoduleerde ruis.

Uitgangspunt bij het bepalen van de STI vormt een speciaal testsignaal namelijk ruis waarvan de intensiteit met een cosinusfunctie is gemoduleerd. De intensiteit van het signaal is te schrijven als:

$$I_i(t) = \bar{I}_i(1 + \cos 2 \pi Ft) \quad (1)$$

Met:

\bar{I}_i tijdgemiddelde intensiteit in W/m^2
F modulatiefrequentie in Hz

Wanneer het ruisspectrum op een bepaalde manier wordt gewogen en de modulatiefrequentie in het gebied tussen 0 en 20 Hz wordt gekozen, is dit testsignaal representatief voor lopende spraak, waarin immers de lettergrepen elkaar met frequenties van deze orde opvolgen. Wanneer het signaal zich voortplant door een ruimte met reflecties, galm en stoorlawaai neemt de modulatie diepte van de intensiteit af; op een ontvangerpositie is de intensiteit van het daar waargenomen ruissignaal te schrijven als

$$I_o(t) = \bar{I}_o (1 + m \cos 2\pi F (t - \theta)) \quad (2)$$

Met:

- m resulterende modulatie diepte
 θ faseverschuiving in het modulatiefrequentiedomein in graden

m hangt af van F; de curve die het verband aangeeft tussen m en F staat bekend als modulatieoverdrachtsfunctie ("Modulation Transfer Function", MTF).

Houtgast en Steenekens gebruikten de MTF als basis van de "Speech Transmission Index" (STI) die een goede maat blijkt te zijn voor de spraakverstaanbaarheid in een ruimte. Hiertoe wordt de MTF bepaald voor verschillende octaafbanden van het spraakspectrum (meestal wordt volstaan met de octaafband van 500 Hz en 2000 Hz en "vertaald" in een equivalente signaalruisverhouding:

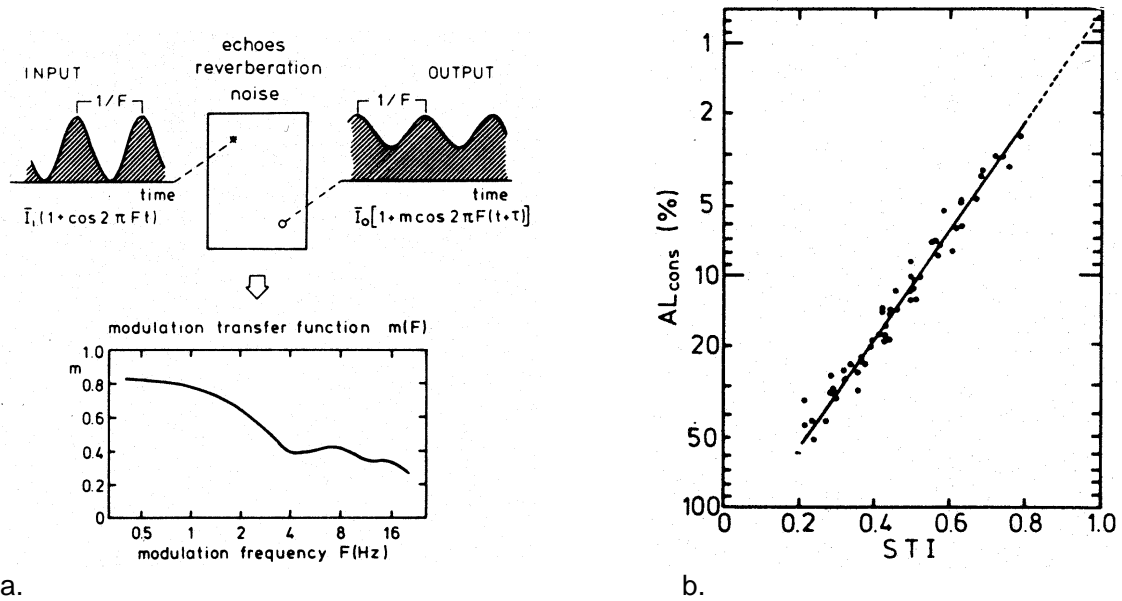
$$S / R_{eq}(F) = 10 \log \frac{m(F)}{1 - m(F)} [dB] \quad (3)$$

Vervolgens worden waarden van $S/R_{eq}(F)$ die groter dan +15 dB of kleiner dan -15 dB zijn "geclippt", waarna S/R_{eq} over de beschouwde F-waarden wordt gemiddeld. De zo voor diverse octaafbanden gevonden waarden worden door gewogen optelling gecombineerd tot een "overall-gemiddelde: waarde $\overline{S / R_{eq}}$.

Ten slotte wordt STI gevonden door $\overline{S / R_{eq}}$ zodanig te schalen dat de uitkomst tussen 0 en 1 ligt; dit via de formule:

$$STI = \frac{\overline{S / R_{eq}} + 15}{30}. \quad (4)$$

Figuur 1a illustreert bovenstaande procedure, terwijl figuur 1b laat zien hoe STI-waarden corresponderen met de "PB-word score" dat wil zeggen het door proefpersonen juist verstaane percentage van een lijst met nonsens-lettergrepen waarin rekening is gehouden met de regelmaat waarin bepaalde letters in onze taal voorkomen; "e" vaker dan "z", enz. ("Phonetically Balanced", PB).



figuur 1. Speech Transmission Index; a = methodiek; b. = STI en spraakverstaanbaarheid

STI	Spraakverstaanbaarheid
<0,4	Slecht
0,4 - 0,6	matig
0,6 - 0,8	goed
0,8 – 1,0	uitstekend

tabel 1. kwalitatieve beoordeling van de spraakverstaanbaarheid als functie van de STI-waarde

Tabel 1 geeft aan hoe de STI-waarde correspondeert met de kwaliteit van de spraakverstaanbaarheid. Voor een goede spraakverstaanbaarheid moet de STI-waarde minstens 0,6 bereiken.

3 Articulation Loss of Consonants

Een andere methode is ontwikkeld door Peutz en Klein [1]. Zij stelden als maat voor de spraakverstaanbaarheid voor de "Articulation Loss of Consonants" ($AL_{cons.}$). Dit is het percentage fout verstaane medeklinkers in fonetisch gebalanceerde klinker-medeklinker-klinker woorden, uitgesproken in draagzinnen. Voor een bij benadering ruisvrij spraaksignaal hangt AL_{cons} af van de nagalmtijd T en de direct-galm-verhouding, namelijk:

$$AL_{cons} \approx 0,73 \frac{T}{r_g^2 / r^2} [\%] \quad \text{voor } r \leq 3,5 r_g; \quad (5)$$

$$AL_{cons} \approx 9T [\%] \quad \text{voor } r > 3,5 r_g. \quad (6)$$

Met:

- AL_{cons} articulation loss of consonants
- T nagalmtijd in s
- r afstand spreker en ontvanger in m
- r_g galmstraal in m (grenswaarde tussen direct en diffuus geluidveld)

Stijgende waarden van AL_{cons} betekenen een afname van de spraakverstaanbaarheid. Voor luisterplaatsen “ver” in het galmveld, waar de afstand r tot de bron meer dan 3,5 maal de galmstraat r_g bedraagt, neemt AL_{cons} met de nagalmtijd evenredig toe; de verstaanbaarheid neemt dus af bij langere nagalmtijd, zoals ook de ervaring leert. Voor afstanden r waar ook het directe geluid een rol speelt, neemt AL_{cons} evenredig met de direct-galm-verhouding af; de spraakverstaanbaarheid neemt dus, zoals te verwachten, toe wanneer men dichterbij de spreker gaat zitten.

De betekenis van AL_{cons} voor de spraakverstaanbaarheid is:

$AL_{\text{cons}} \approx 15\%$:	ontstaan van klachten over de spraakverstaanbaarheid tenzij slechts eenvoudige informatieoverdracht plaatsheeft;
$10\% \leq AL_{\text{cons}} < 15\%$	slechte sprekers leiden respectievelijk slechte luisteraars hebben klachten bij ingewikkelde boodschappen;
$AL_{\text{cons}} < 10\%$	goede spraakverstaanbaarheid.

4 Speech Interference Level (SIL)

Ook het niet realiseren van spraakverstaanbaarheid, dat wil zeggen het creëren van een akoestische omgeving waarin de inhoud van een gesprek vertrouwelijk blijft, kan de doelstelling van bouwakoestische maatregelen zijn. Te denken hierbij is aan grotere kantoorruimte of een balieruimte in een publiekshal.

Maat voor de verstaanbaarheid van een gesprek in een akoestische omgeving is de Speech Interference Level (SIL). Het verschil met STI en AL_{cons} is dat bij de SIL het niveau niet van de spraakbron afkomstig is maar van “vreemde” bronnen. De SIL is derhalve een maat voor de invloed van achtergrondgeluid op de spraakverstaanbaarheid. Basis is de gemiddelde waarde van het achtergrondniveau in de 500, 1000, 2000 en 4000 Hz octaafband.

Achtergrondniveaus met dezelfde SIL-waarde worden geacht hetzelfde effect op de spraakverstaanbaarheid te hebben. De SIL kan worden gebruikt om de maximaal toelaatbare afstand tussen spreker en luisteraar (face-to-face communication) te bepalen waarbij deze een acceptabele spraakverstaanbaarheid aanwezig is (zie tabel 2).

afstand in m spreker/ontvanger	stemniveau			
	normaal	verhoogd	zeer luid	schreeuwen
0,3	65	71	77	83
0,9	55	61	67	73
1,5	51	57	63	69
1,8	49	55	61	67
3,7	43	49	55	61

tabel 2. maximaal toelaatbare SIL-waarde in [dB] van stoorgeluid voor het verkrijgen van een goede spraakverstaanbaarheid

In het algemeen geldt voor een vrouwelijke spreker een reductie van de SIL-waarde van 5 dB.

5 Articulation index

De hiervoor behandelde SIL-methode houdt slechts beperkt rekening met de frequentiesamenstelling van het stoorgeluid; immers, de SIL-waarde is een gemiddelde waarde over 4 octaafbandniveaus.

Voor de spraakinformatieoverdracht is het frequentiegebied van de 2000 Hz en 4000 Hz octaafband het belangrijkste (zie figuur 2) waarin als functie van de octaafband de procentuele bijdrage van een spraaksignaal aan de spraakverstaanbaarheid door een luisteraar is weergegeven.

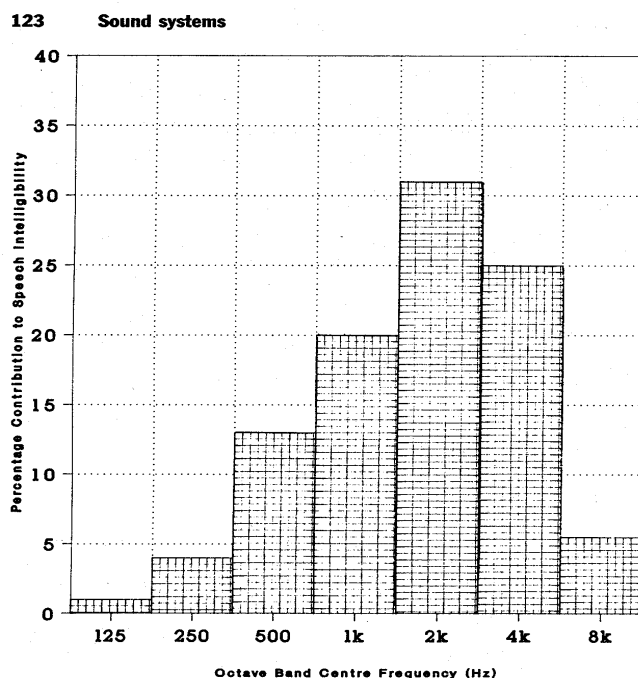
De articulation index houdt hier rekening mee door de signaal-ruis-verhouding per tertsband als uitgangspunt te nemen. De methode is al in 1969 genormaliseerd (ANSI 53.5).

De signaal-ruis-verhouding van het spraaksignaal en het stoorgeluid wordt voor ieder van de 20 tertsbanden (middenfrequentie van 200 Hz – 5 kHz) bepaald en vervolgens gewogen; de weegfactoren zijn gegeven in tabel 3. De som van de gewogen signaal-ruis-verhouding geeft de articulation index. De AI varieert tussen 0 en 1 waarbij AI = 1 betekent dat er een 100% spraakverstaanbaarheid aanwezig is.

De relatie tussen AI en de spraakverstaanbaarheid is gegeven in figuur 3, globaal betekent dit:

AI \leq 0,2	niet acceptabel
AI = 0,2 – 0,3	marginale spraakverstaanbaarheid
0,3 – 0,4	acceptabel
0,4 – 0,5	goed
0,5 – 0,6	zeer goed
0,6 – 0,7	uitstekend

Van secundair belang voor de waardering van AI is het type informatie dat wordt overgedragen, de snelheid van de spraak enz..

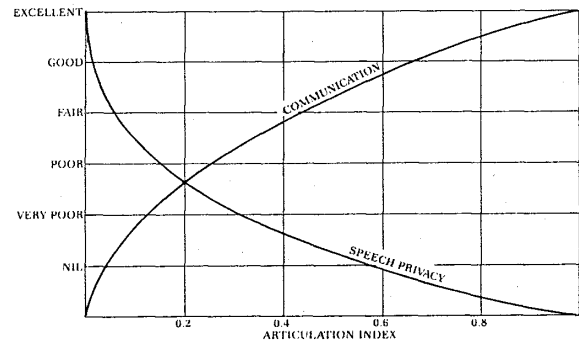


figuur 2. bijdrage per octaafband aan de spraakverstaanbaarheid

Articulation index calculation table

1/3 OBCF (Hz)	Speech peaks minus noise (dB)	Weight	Contribution to articulation index
200		0.000 4	
250		0.001 0	
315		0.001 0	
400		0.001 4	
500		0.001 4	
630		0.002 0	
800		0.002 0	
1 K		0.002 4	
1.25 K		0.003 0	
1.6 K		0.003 7	
2 K		0.003 7	
2.5 K		0.003 4	
3.15 K		0.003 4	
4 K		0.002 4	
5 K		0.002 0	

Relationship between articulation index, Communication and speech privacy



tabel 3.

figuur 3.

Zoals in het bovenstaande al naar voren komt, hangt de spraakverstaanbaarheid af van parameters als nagalmtijd, galmstraal en signaal-ruis-verhouding. In het algemeen kan worden gesteld dat voor een goede spraakverstaanbaarheid een sterk direct geluid nodig is, zo mogelijk ondersteund door enkele vroege reflecties ($\tau \leq 15$ ms); "sterk" heeft hier betrekking op de energieverhouding ten opzichte van late reflecties en galm.

Bij dit alles moet een laag achtergrondniveau zorgen voor een voldoende waarde van de signaal-ruis-verhouding; in spreekzalen wordt dikwijls de eis gesteld dat het achtergrondniveau de 25 dB(A) niet overschrijdt. De aanwezigheid van echo's (dat wil zeggen sterke, late reflecties die in de pulsresponsie boven hun "omgeving" uitsteken en als een herhaling van het directe geluid hoorbaar zijn) is voor spraakverstaanbaarheid desastreus en moet tot elke prijs worden vermeden.

Optimale overdracht van het directe geluid is gewaarborgd wanneer er goede zichtlijnen van luisteraar naar spreker aanwezig zijn: goed zien is een vereiste voor goed horen! Het is daarom voor de spraakverstaanbaarheid - en even zo goed voor de perceptie van muziek - gunstig om de stoelen in een zaal vanaf het podium olopend te rangschikken (tribune, amfitheater), zoals reeds de Grieken en Romeinen deden in hun openluchttheaters.

Klankkaatsers achter of boven de spreker - zoals bij kansels in vele kerken - genereren vroege reflecties die door het gehoormechanisme van de luisteraar geïntegreerd met het directe geluid worden waargenomen en worden ervaren als een versterking van het directe geluid. Ze dragen dus effectief bij aan verhoging van de direct-galm-verhouding. Vroege reflecties kunnen ook langs elektronische weg worden opgewekt.

Verdere verhoging van de direct-galm-verhouding wordt bereikt door de nagalmtijd van de ruimte te reduceren door de afmetingen, met name de hoogte, niet groter te maken dan noodzakelijk en eventueel extra absorptiemateriaal op wanden en plafond aan te brengen. De meest letterlijke vorm van "open raam" vindt men in eerder genoemde antieke openluchttheaters, waar galm geheel ontbreekt. Zeker wanneer een "klankmuur" voor vroege

reflecties zorgt en het achtergrondniveau (ook nu nog!) laag is, is de spraakverstaanbaarheid in zulke theaters vaak zo goed dat elektronische versterking overbodig is.

Uit het bovenstaande zou men kunnen concluderen dat in een spreekzaal nagalm alleen maar nadelig is. Voor het spraakverstaan door de luisteraar is dit inderdaad het geval; echter, niet voor het "spreekgemak" van de spreker. Voor hem is van belang dat zijn woorden ook letterlijk enige "weerklink" in de zaal vinden; anders krijgt hij onbewust het gevoel dat zijn woorden niet worden verstaan (ten onrechte), gaat dus harder praten en forceert zijn stem. Daarom wordt als geschikte nagalmtijd in een spreekzaal een waarde tussen 0,5 en 1,0 seconde aangehouden.

LITERATUUR

1. V. Peutz, Klein; "Spraakverstaanbaarheid onder gecombineerde invloed van ruis, nagalm en echo; NAG-publicatie 28.