

De fysica van de spraakverstaanbaarheid

Ir. T. HOUTGAST

Instituut voor Zintuigfysiologie RVO-TNO, Soesterberg

The physics of intelligibility

Samenvatting

De kwaliteit van de communicatieweg tussen een spreker en een luisteraar is bepalend voor de verstaanbaarheid van de spraak zoals de luisteraar die hoort. Een nieuwe methode om deze kwaliteit te meten berust op resultaten van fundamentele onderzoeken betreffende de waarneming van klanken. Daarbij bleek dat de verschillen tussen klanken zoals een luisteraar die ervaart, goed kwantificeerbaar samenhangen met de verschillen in de bijbehorende frequentiespectra. Dit suggereert dat de variaties in het frequentiespectrum van de spraak bepalend zijn voor de verstaanbaarheid. Deze variaties kunnen gekwantificeerd worden als de som van de varianties van de geluidsdruk-niveaus in achttien frequentiebanden met een breedte van $1/3$ oktaaf, de spectrale variantie. De kwaliteit van een communicatieweg wordt dan bepaald door de mate waarin de oorspronkelijk in de spraak aanwezige spectrale variantie aan de luisterzijde bewaard is gebleven. Door gebruik te maken van een specifiek testsignaal kan de „variantiereductie” door de communicatieweg, en daarmee de invloed op de spraakverstaanbaarheid op eenvoudige wijze worden gemeten.

Een experiment waarbij deze methode gebruikt werd naast de traditionele en tijdrovende methode met sprekers en luisteraars laat een goede correlatie zien tussen de uitkomsten van beide methoden.

Inleiding

Wanneer de spreker in de zaal vraagt „kunt U mij achterin ook verstaan?”, of de radiotelegrafist zegt „ontvangst sterk gestoord, sterkte 1 à 2”, dan zijn dit symptomen van het algemene probleem dat de spraak, op z'n weg van spreker naar luisteraar zodanig gestoord kan worden, dat de verstaanbaarheid voor de luisteraar een probleem wordt. Deze storingen kunnen van velerlei aard zijn, zoals maskering door lawaai, bandbreedtebeperking, vervorming, nagalm. Van elk dezer storingen is wel ongeveer bekend in hoeverre ze toelaatbaar zijn in verband met de verstaanbaarheid; hiermee houdt de ontwerper, zoals de zaalakoesticus of de elektronicus natuurlijk terdege rekening.

Naast het hanteren van dergelijke vuistregels bestaat er bij het ontwerpen en testen van communicatiewegen behoefte aan een *meetmethode* waarmee de kwaliteit bepaald kan worden, bijvoorbeeld om een inzicht te krijgen betreffende de invloed van een klankkaatser boven een spreekge-

Summary

A new approach in measuring the intelligibility in speech communication rests on results of basic experiments on the perception of sounds. It was found that differences in frequency spectra of sounds are directly related to perceptual differences and, as a consequence, it was assumed that the variations in the frequency spectrum of speech are directly related to intelligibility. These variations are mathematically quantified by the sum of the variances of the sound pressure levels in eighteen frequency bands with a width of $1/3$ octave, the spectral variance. Thus a direct measure of the intelligibility might be obtained by analyzing the speech signal as heard by the listener and measuring the spectral variance. This analysis can be facilitated by using a specific test signal instead of a speech signal.

An experiment concerning speech intelligibility in which this method is used as well as the traditional method with several speakers and listeners shows a good correlation between the results of both methods.

stoelte of het verlengen van de antenne van een zend-ontvanger.

Hiertoe staan in het algemeen twee wegen open. De eerste methode sluit zo goed mogelijk aan bij het gebruik: men laat een aantal sprekers iets zeggen en men gaat met een aantal luisteraars na hoe de verstaanbaarheid is. Deze subjectieve methode is voor het verkrijgen van betrouwbare resultaten zeer tijdrovend, doch voor een onderzoek van beperkte omvang goed te gebruiken. Bij de andere methode tracht men langs fysische weg de optredende storingen te kwantificeren en via de bekende vuistregels te interpreteren in termen van verstaanbaarheid. Hoewel deze methode snel kan zijn, blijft de interpretatie van de storingen, speciaal wanneer verschillende storingen tegelijk optreden, een zwakke schakel in deze procedure.

Steunend op nieuwe inzichten, verkregen uit algemene fundamentele onderzoeken betreffende de waarneming van geluiden, is getracht een nieuwe

methode te ontwikkelen die een synthese vormt van het *objectieve* en *snelle* van de fysische methode en de *betrouwbaarheid* van de directe benadering volgens de subjectieve methode.

Analyse

Een nadere analyse van de beide methoden geeft aan wat het uitgangspunt dient te zijn. Bij de methode spreker-luisteraar wordt de kwaliteit van het ontvangen spraaksignaal direct beoordeeld, zonder te letten op de aard van de eventuele storingen. Bij de fysische methode worden juist deze storingen gemeten en wordt dus *langs een omweg* de kwaliteit van het te ontvangen signaal bepaald. Deze omweg kunnen we kortsluiten indien we in staat zijn om langs fysische weg het ontvangen signaal zelf te beoordelen op een wijze die direct aansluit bij de verstaanbaarheid. De verstaanbaarheid is direct gekoppeld met de onderscheidbaarheid van de klanken die in de spraak voorkomen. Het gaat dus om het vinden van een fysische grootheid in het ontvangen spraaksignaal die eenduidig gekoppeld is met de onderscheidbaarheid van de diverse spraakklanken door de luisteraar, dus met de *perceptieve verschillen* tussen de spraakklanken.

Gegevens uit fundamentele onderzoeken

Voor deze benadering kunnen we steunen op de resultaten van experimenten waarbij onderzocht is hoe de perceptieve eigenschappen van klanken gekoppeld zijn met de fysische eigenschappen [1]. Daarbij werd gevonden dat het perceptieve verschil tussen twee klanken goed kwantificeerbaar samenhangt met de verschillen in de frequentiespectra. Een vruchtbare voorstellingswijze hiervan wordt als volgt opgebouwd. Het frequentiespectrum van een klank kan worden bepaald door de geluiddrukkniveaus (in dB) te meten in 18 aaneengesloten frequentiebanden ter breedte van $\frac{1}{3}$ oktaaf, het gehele gebied van de lage tot de hoge frequenties bestrijkend. (De bandbreedte van $\frac{1}{3}$ oktaaf sluit goed aan bij de filtereigenschappen van het gehoororgaan). Door deze 18 getallen is het spectrum van de klank dus vastgelegd. Men kan nu deze 18 getallen beschouwen als coördinaten langs 18 onderling loodrechte assen, waardoor de klank voorgesteld is als een punt in deze 18-dimensionale ruimte. Men kan deze ruimte de *spectrale ruimte* noemen omdat immers elk geluid via het bijbehorende geluidsspectrum in deze ruimte als een punt kan worden weergegeven. De voor onze benadering essentiële betekenis van deze voorstellingswijze is nu dat de *afstand* tussen twee klanken in deze spectrale ruimte eenduidig gekoppeld blijkt te zijn met het *perceptieve verschil* tussen de twee klanken, zoals gewaardeerd door luisteraars. De onderlinge afstand is een fysische grootheid, bepaald uit fysische metingen, uitgedrukt in

dB. Klanken met een grote afstand in de spectrale ruimte zijn klanken met een groot perceptief verschil, dus klanken die zich duidelijk van elkaar onderscheiden, terwijl klanken met een kleine onderlinge afstand moeilijk van elkaar te onderscheiden zijn.

De onderlinge afstand in deze spectrale ruimte dient dus het fysisch uitgangspunt te zijn om iets te zeggen omtrent de onderscheidbaarheid van klanken en tenslotte over de verstaanbaarheid.

Spectrale variantie en onderscheidbaarheid

Een groep klanken kan in de spectrale ruimte voorgesteld worden als een aantal punten, een puntenconfiguratie. Wanneer enigerlei storing wordt geïntroduceerd, dan worden de posities van de punten in de spectrale ruimte bepaald door de spectra van de gestoorde klanken. De oriëntatie van de puntenconfiguratie is dus sterk afhankelijk van de aard van de storing. Gezien het voorgaande mag men veronderstellen dat de onderscheidbaarheid van de klanken binnen die groep gekoppeld is met de uitgebreidheid van de bijbehorende puntenconfiguratie; immers met een grote uitgebreidheid gaan grote onderlinge afstanden gepaard en daarmee grote perceptieve verschillen. Een wiskundige maat voor de uitgebreidheid is de *variantie* van de puntenconfiguratie, die berekend wordt als het gemiddelde van de kwadraten van alle onderlinge afstanden binnen de configuratie. We noemen dit de spectrale variantie, omdat het gebaseerd is op de afstanden in de spectrale ruimte. Deze veronderstelling, dat de onderscheidbaarheid van een groep klanken uitsluitend bepaald wordt door de spectrale variantie van de overeenkomstige puntenconfiguratie in de spectrale ruimte, ongeacht de specifieke aard van de storing, is experimenteel onderzocht. De groep klanken voor dit experiment bestond uit de zeven klinkers ie, o, u, a, e, i, oe, uitgesproken in de woorden h (klinker) t. Dit signaal kon aan twee typen storing onderworpen worden: toevoegen van ruis (sterkte instelbaar) en bandbreedtebeperking (vier bandfilters).

Gemeten werd bij veertien condities, bestaande uit verschillende combinaties van achtergrondlawaai en filtering. Enerzijds werd bij elk van deze condities de verstaanbaarheid gemeten door de zeven testwoorden vele malen in willekeurige volgorde aan negen luisteraars te laten horen en het over deze luisteraars gemiddelde percentage correct genoteerde woorden te bepalen.

Anderzijds werd bij elk van deze condities, door nauwkeurige bepaling van de frequentiespectra van de zeven gestoorde klinkers zoals gehoord door de luisteraars, de spectrale variantie bepaald. De overeenkomstige waarden van de verstaanbaarheid (%) en de spectrale variantie (dB)² zijn in Fig. 1 weergegeven.

Het blijkt dat de spectrale variantie inderdaad een

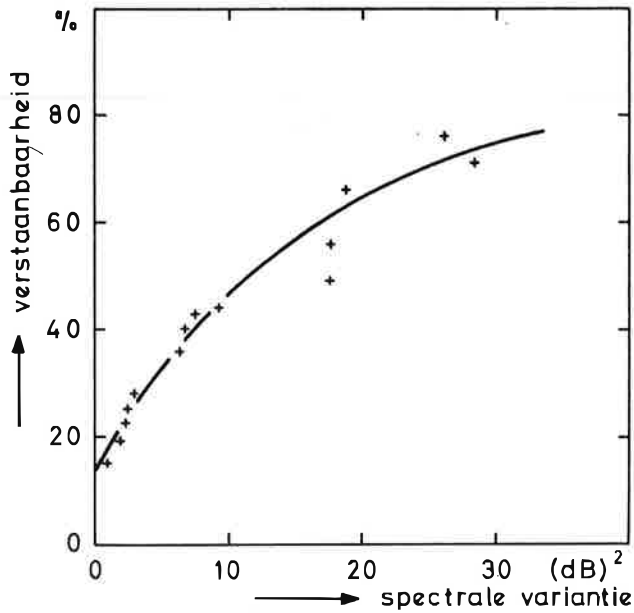


Fig. 1. Bij veertien combinaties van verschillende bandfilters en verschillende sterkten van het stoorlawaai werd de onderscheidbaarheid van een groep van zeven klinkers onderzocht. De perceptieve onderscheidbaarheid (% correct genoteerd door luisteraars) is uitgezet tegen de fysische onderscheidbaarheid (spectrale variantie op grond van de spectra van de gestoorde klinkers).

goede maat is voor de verstaanbaarheid, onafhankelijk van de specifieke aard van de storing. Het meten van de verstaanbaarheid van een groep klinkers kan dus vervangen worden door het langs fysische weg bepalen van de spectrale variantie van de groep klinkers.

In principe

Het langs fysische weg bepalen van de verstaanbaarheid van een spraaksignaal kan nu in principe als volgt gebeuren. De spraak wordt, zeer schema-

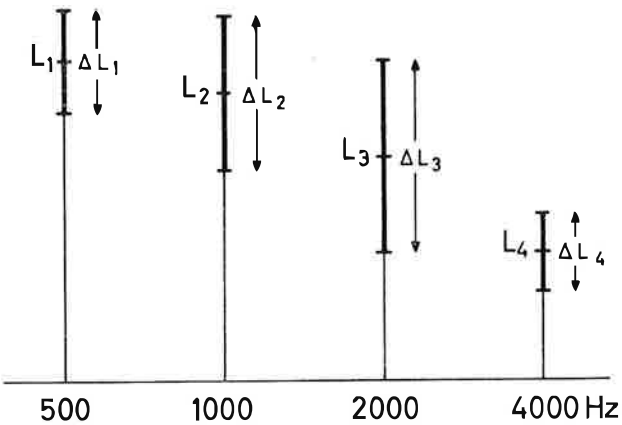


Fig. 2. Het frequentiespectrum van het testsignaal. Elk der vier frequentiecomponenten is in sterkte gemoduleerd in een ritme van 10 Hz. De modulatie diepte ΔL en de gemiddelde sterkte L zijn aangepast aan de statistische eigenschappen van de spraak.

tisch, opgevat als een voortdurende opeenvolging van klinken met verschillende spectra, in ruwe benadering 10 klinken per seconde. De gestoorde spraak, zoals de luisteraar die hoort, wordt nu opgedeeld in een opeenvolging van bemonsteringen met een duur van 100 ms.

Van elk daarvan wordt het spectrum bepaald en op grond daarvan de bijbehorende positie in de spectrale ruimte. De opeenvolging van bemonsteringen resulteert in een puntenwolk in de spectrale ruimte, waarvan de uitgebreidheid, de variantie, bepaald kan worden. In het licht van het voorgaande mag men verwachten dat de op deze wijze bepaalde variantie van de gestoorde spraak een goede maat is voor de verstaanbaarheid.

De praktische moeilijkheid doet zich voor dat eventuele fluctuaties van een stoorsignaal, dat gesuperponeerd is op het spraaksignaal, bij deze procedure bijdraagt tot de te meten variantie, waardoor een verkeerde indicatie betreffende verstaanbaarheid verkregen wordt. Men dient dus de relevante fluctuaties (van de spraak) te scheiden van de irrelevante fluctuaties (van de storing).

Praktische uitvoering

Dit probleem wordt opgelost door niet de spraak als signaal te gebruiken, maar een zeer specifiek testsignaal, waarbij de geïntroduceerde fluctuaties in het frequentiespectrum een voor een apparaat „herkenbaar” karakter hebben, namelijk een ritme van exact tien variaties per seconde. Aan de ontvangtzijde kunnen nu, door middel van een bandfilter van 10 Hz, deze relevante fluctuaties gescheiden worden van de irrelevante fluctuaties in een stoorsignaal, die over het algemeen een ander ritme dan exact 10 Hz zullen hebben.

De verdere eigenschappen van dit testsignaal dienen natuurlijk goed aan te sluiten bij de statistische eigenschappen van de normale spraak; anderzijds dient men zich te hoeden voor onnodige detaillering. Een bruikbaar testsignaal is weergegeven in

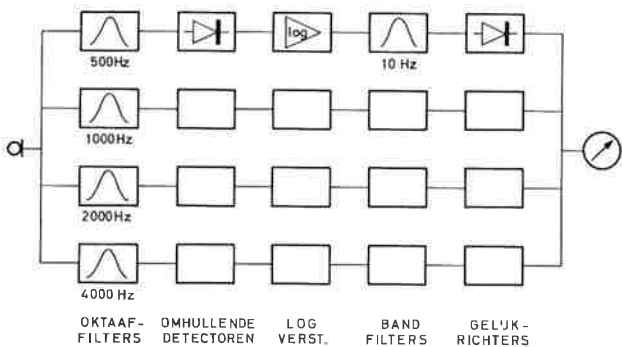


Fig. 3. Blokschema van het analyse-apparaat. Van het ontvangen signaal wordt voor elk der vier frequentiecomponenten bepaald in welke mate de oorspronkelijke sterktemodulatie met het ritme van 10 Hz bewaard is gebleven, hetgeen bepalend is voor de kwaliteit van de betreffende communicatieweg.

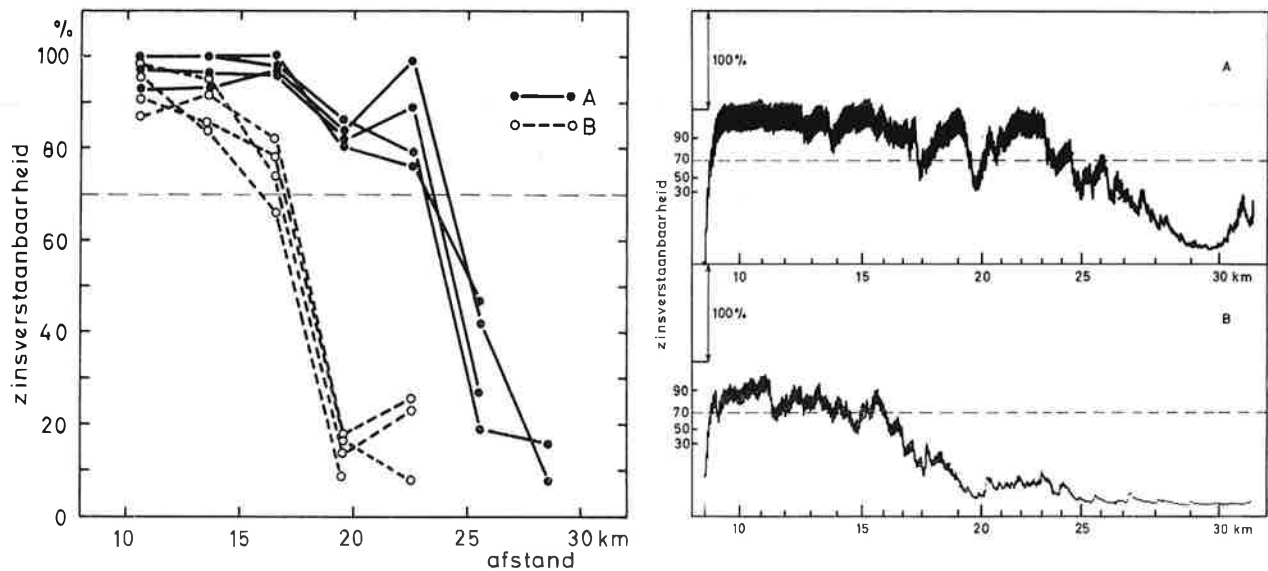


Fig. 4. De spraakverstaanbaarheid van een verbinding tussen een zender en ontvanger als functie van de onderlinge afstand. Twee typen zend-ontvangers (A en B).
 Links: resultaat van de tijdrovende methode met sprekers en luisteraars.
 Rechts: resultaat van de methode met testsignaal en analyse-apparatuur.

Fig. 2. Het is samengesteld uit slechts vier frequentiecomponenten: 500, 1000, 2000 en 4000 Hz en kan dus niet het gedetailleerde beeld geven dat gebaseerd is op een analyse in 18 frequentiebanden. Hier is slechts een analyse zinvol in vier frequentiebanden, namelijk de oktaafbanden met middenfrequenties van 500, 1000, 2000 en 4000 Hz. Gezien de aard van de meeste storingen mag verwacht worden dat het informatieverlies ondanks deze sterke vereenvoudiging toch gering is. Elk van deze frequentiecomponenten is in sterkte gemoduleerd met het ritme van 10 Hz, waarbij het gemiddelde en de modulatie diepte voor elke frequentiecomponent als volgt bepaald zijn:

- De gemiddelde geluidrukniveaus van de vier frequentiecomponenten, L_1 tot L_4 in Fig. 2 komen overeen met de gemiddelde geluidrukniveaus van de normale spraak in de betreffende oktaafbanden.
- De variaties om deze gemiddelde waarden, ΔL_1 tot ΔL_4 in Fig. 2 worden bepaald door de variaties van het geluidrukniveau van de normale spraak binnen de oktaafbanden. De som $\Delta L_1 + \Delta L_2 + \Delta L_3 + \Delta L_4$ wordt dus bepaald door de totale spectrale variantie die in de normale spraak aanwezig is.

Dit testsignaal doorloopt geheel dezelfde weg als het spraaksignaal zou doorlopen. Van het ontvangen testsignaal wordt nu, met behulp van het in Fig. 3 weergegeven meetapparaat, nagegaan in hoeverre de oorspronkelijk aanwezige variaties in de vier frequentiebanden bewaard zijn gebleven. Het meetapparaat geeft de som van deze bewaard gebleven variaties, hetgeen een directe maat is voor de spectrale variantie in een spraaksignaal zoals

dat door een luisteraar via deze communicatieweg gehoord zal worden, en daarmee voor de verstaanbaarheid. De uitgangsspanning van dit meetinstrument is dus een directe maat voor de kwaliteit van de betreffende communicatieweg.

Een toepassing

Het zal duidelijk zijn dat in het voorgaande slechts de hoofdlijnen zijn weergegeven en dat vele details niet aan de orde zijn gekomen. Bij de hier te vermelden toepassing wijkt het testsignaal zowel als de analyseapparatuur dan ook enigszins af van wat hiervoor geschetst werd, maar het principe blijft dat gemeten wordt hoe groot de spectrale variantie van een spraaksignaal zal zijn dat via de betreffende communicatieweg de luisteraar bereikt. Van twee typen zend-ontvangers (type A en B) diende nagegaan te worden hoe de kwaliteit van de verbinding was als functie van de onderlinge afstand van de zender en ontvanger. Hiertoe was de zender gemonteerd in een rijdend voertuig. De afstand tussen zender en ontvanger werd continu gevarieerd tussen 10 en 30 km.

Het resultaat van de metingen volgens de methode spreker-luisteraar staat weergegeven in Fig. 4, links. Voor elk type werd de meting vier maal verricht, waarbij in totaal 12 personen betrokken waren. Deze metingen waren zeer tijdrovend (meer dan 100 man-uren). Om een indruk te krijgen van de nieuwe meetmethode werd ook deze toegepast. Het spraaksignaal werd vervangen door het testsignaal en de luisteraar door de analyse-apparatuur. Deze metingen namen slechts enkele uren in beslag.

Het resultaat staat ook in Fig. 4 weergegeven, ge-

scheiden naar type zend-ontvanger. Hier is continu de uitgangsspanning van de analyse-apparatuur uitgeschreven tijdens het vergroten van de afstand tussen zender en ontvanger. Deze uitgangsspanning, geïjkt in % zinsverstaanbaarheid, blijkt inderdaad goed overeen te komen met de verstaanbaarheid zoals gemeten met sprekers en luisteraars.

Samenvattend

Als we in het kort de gedachtengang nog eens weergeven die leidt tot het langs fysische weg bepalen van de spraakverstaanbaarheid dan wordt het uitgangspunt gevormd door het beeld van de spectrale ruimte, zoals dat gegroeid is uit de resultaten van fundamentele onderzoeken. Klanken vinden op grond van hun spectra een plaats in deze veel-dimensionale ruimte, waarbij de onderlinge afstanden een maat zijn voor de perceptieve verschillen tussen de klanken, zoals waargenomen door luisteraars. Dit legt dus het verband tussen de perceptie en de fysica. Voortbouwend hierop kan

de normale spraak, opgebouwd uit een opeenvolging van vele klanken, op grond van de bijbehorende frequentiespectra als een puntenwolk in deze ruimte worden voorgesteld. De oriëntatie van deze puntenwolk in de spectrale ruimte wordt sterk beïnvloed door eventuele storingen. Het blijkt dat de uitgebreidheid van de puntenwolk, de spectrale variantie, het fysisch equivalent is voor de verstaanbaarheid, ongeacht de specifieke aard van de storing.

Als de praktische moeilijkheden bij het bepalen van de spectrale variantie van de gestoorde spraak zoals de luisteraar die hoort, zijn opgelost, dan kan, op grond van deze spectrale variantie langs fysische weg de kwaliteit van een verbinding tussen spreker en luisteraar bepaald worden.

Literatuur

- [1] L. C. W. Pols, L. J. Th. v. d. Kamp and R. Plomp: Perceptual and Physical Space of Vowel Sounds. Ter publikatie aangeboden aan The Journal of the Acoustical Society of America