

in which way spectral contrasts are affected by the reduction of temporal modulations. As a counterpart of the MTF for measuring the transfer of temporal modulations, a spectral modulation transfer function (SMTF) is presented to determine the extent to which modulations (ripples per octave) in the short-term spectral envelope are preserved. Uniform reduction (i.e., independent of modulation frequency) of the temporal modulations in $\frac{1}{4}$ -oct bands appears to result in a virtually equal degree of reduction of the spectral modulations. Measurements of the SMTF after temporal-envelope filtering (chapters 2 and 3) were carried out for a number of lowpass and highpass cutoff frequencies (1-32 Hz). When the effect of this type of temporal processing is expressed in terms of the associated reduction of spectral contrasts, an interesting question is whether spectral reduction alone can explain the deleterious effect on intelligibility. Sentences were processed with an analysis-resynthesis algorithm consisting of short-term fast Fourier transforms, 10%-90% contrast reduction in the spectral envelope, inverse transforms, and overlapping additions to reconstruct a continuous signal. The SRT in noise for the spectrally reduced sentences was measured with 10 subjects. Comparison of these SRTs with those obtained in chapters 2 and 3 reveals that the loss of intelligibility due to temporal highpass filtering can be attributed to the associated spectral modulation reduction. This result is not found for temporal lowpass filtering, suggesting that in that case spectral reduction is only a secondary effect.

Samenvatting

Spraak kan worden opgevat als een verzameling opeenvolgende (smalle) frequentiebanden met amplitude-gemoduleerde signalen. Iedere frequentieband bestaat uit een fijnstructuur (draaggolf) en een in de tijd variërende omhullende. Deze temporele omhullende geeft dus voor een beperkt frequentiegebied de variaties in geluidsterkte weer. Daarin ligt de informatie besloten die essentieel is voor het identificeren van fonemen, lettergrepen, woorden en zinnen. Voor in de praktijk voorkomende verstoringen van de spraak, zoals ruis en nagalm, blijken de nadelige effecten op het verstaan een gevolg te zijn van een vermindering van de temporele modulatie diepte. De mate waarin de temporele omhullende van het spraaksignaal behouden blijft is uitvoerig bestudeerd aan de hand van de modulatie-overdrachtsfunctie (MTF). Deze succesvolle benadering heeft geleid tot de ontwikkeling van een fysische maat voor de kwaliteit van spraakoverdracht, de spraaktransmissie index (STI). Het doel van dit onderzoek is op een systematische manier na te gaan in hoeverre temporele modulaties bepalend zijn voor de verstaanbaarheid. Anders gezegd, wat is het effect van vermindering van de temporele modulaties in het spraak-signaal? Het onderzoek is primair fundamenteel van aard, maar heeft ook een meer toegepaste kant. Kennis over de mate waarin de perceptie van spraak bestand is tegen degradatie levert gegevens op die bijvoorbeeld kunnen worden toegepast in signaalbewerkingen t.b.v. slechthorenden.

Voor de signaalbewerking is een algemeen analyse-resynthese algoritme ontwikkeld, waarmee de temporele (Hilbert) omhullende van een aantal opeenvolgende frequentiebanden kan worden gewijzigd. Op die manier is het mogelijk de omhullende te manipuleren zonder de fijnstructuur aan te tasten. Het effect van deze manipulaties op de verstaanbaarheid van eenvoudige zinnen is gemeten bij jonge normaalhorende luisteraars. Afhankelijk van de experimentele conditie werd het percentage correct verstane zinnen in stilte bepaald of werd de zogenaamde spraakverstaanbaarheidsdrempel (SRT) tegen een achtergrond van ruis gemeten. De SRT is gedefinieerd als die spraak-ruisverhouding in dB waarbij nog 50% van de zinnen foutloos kan worden gereproduceerd. De ruis heeft daarbij hetzelfde gemiddelde frequentiespectrum als de zinnen.

Na een algemene inleiding in hoofdstuk 1 richten de hoofdstukken 2 en 3 zich op de vraag welke omhullende-frequenties van belang zijn voor de verstaanbaarheid. Hoofdstuk 2 beschrijft een serie experimenten waarin de temporele omhullende laagdoorlaat wordt gefilterd (versmeerd), waardoor de hogere modulatiefrequenties worden gereduceerd. Het effect van versmering op de verstaanbaarheid werd gemeten als functie van zowel de afsnij-

frequentie van het laagdoorlaatfilter (0-64 Hz) als de breedte van de frequentiebanden ($\frac{1}{4}$, $\frac{1}{2}$, of 1 octaaf) waarop de bewerking plaatsvond. De resultaten bij 36 luisteraars tonen aan dat de verstaanbaarheid voor zeer lage afsnijfrequenties (0-2 Hz) al in stilte ernstig vermindert, vooral voor smalle frequentiebanden. Voor afsnijfrequenties van 4 Hz en hoger werd de SRT in ruis gemeten. Daaruit blijkt dat de verstaanbaarheid progressief toeneemt tot een afsnijfrequentie van 16 Hz, onafhankelijk van de bandbreedte. Met andere woorden, modulatiefrequenties boven 16 Hz leveren geen wezenlijke bijdrage aan de verstaanbaarheid van alledaagse zinnen (vooropgesteld dat lagere modulatiefrequenties wel in het signaal aanwezig zijn). De gemeten SRT-waarden voor afsnijfrequenties vanaf 4 Hz blijken goed overeen te komen met voorspellingen op basis van de STI. Voor lagere afsnijfrequenties is de berekende STI tamelijk laag en houdt geen rekening met de waargenomen afhankelijkheid van de bandbreedte.

Als vervolg op de temporele versmering wordt in hoofdstuk 3 het effect van reductie van lage modulatiefrequenties op eenzelfde wijze bestudeerd. Daartoe werd de temporele omhullende hoogdoorlaat gefilterd, met afsnijfrequenties tussen 1 en 128 Hz. De SRT voor zinnen in ruis kon worden gemeten voor afsnijfrequenties tot en met 32 Hz; daarboven werd de zinsverstaanbaarheid in stilte bepaald. De resultaten met 42 luisteraars geven aan dat de verstaanbaarheid voor afsnijfrequenties tot en met 4 Hz gelijk blijft aan die van onbewerkte spraak. Verhoging van de afsnijfrequentie geeft een geleidelijke afname van de verstaanbaarheid, welke vanaf 16 Hz afhankelijk is van de bandbreedte ($\frac{1}{4}$, $\frac{1}{2}$, of 1 octaaf). Voor een aantal experimentele condities werd de modulatiereductie gemeten aan de hand van de MTF en vergeleken met data over meerkanalige amplitude-compressie. De resultaten daarvan doen vermoeden dat compressie, althans bij normaalhorenden, een beperkte achteruitgang van de verstaanbaarheid geeft (ca. 1 dB verhoging van de SRT). Vergelijken we de resultaten van deze experimenten met die uit het voorgaande hoofdstuk, dan blijkt de 'cross-over' modulatiefrequentie—de frequentie die het temporele modulatiespectrum in twee perceptief even belangrijke stukken verdeelt—voor alle drie de bandbreedten bij 8 à 10 Hz te liggen. Dit betekent dat reductie van temporele modulaties of beneden of boven 8-10 Hz bij een kritische spraak-ruisverhouding tot dezelfde (verminderde) verstaanbaarheid leidt.

Naast de verstaanbaarheid van zinnen is in de hoofdstukken 2 en 3 ook gekeken naar het effect van filteren van de omhullende op de herkenning van individuele fonemen. Hiertoe werden identificatiescores in stilte voor klinkers en medeklinkers in nonsens CVC- en VCV-syllaben gemeten bij verschillende afsnijfrequenties van de laag- en hoogdoorlaatfilters (bewerking in $\frac{1}{4}$ -octaaf banden). Als verwacht neemt de herkenning af naarmate de

afsnijfrequentie van laag- en hoogdoorlaatfilter lager resp. hoger is. Gemiddeld hebben medeklinkers meer te lijden dan klinkers. Fouten in de identificatie van klinkers worden voornamelijk veroorzaakt door slechtere herkenning van diftongen en door kort-lang/lang-kort verwarringen. De medeklinkers kunnen verdeeld worden in drie categorieën, nl. plosieven, fricatieven en klinkerachtigen. De meerderheid van de foutieve medeklinkerbenoemingen blijken verwarringen binnen een categorie te zijn, d.w.z. informatie over de wijze van articuleren blijft min of meer behouden. Uitzondering hierop vormen de versmeerde plosieven, die, afhankelijk van de positie in de syllabe, vaak als fricatief of klinkerachtig benoemd worden. Verwarringen die verwacht mogen worden als gevolg van versmering (b.v. /p/-/f/, /t/-/s/, /k/-/x/) treden vooral op bij finale plosieven.

De resultaten van bovengenoemde experimenten geven aan dat er een minimum aan variatie in de temporele omhullende nodig is die zorgt voor voldoende modulatie diepte, d.w.z. voldoende verschil tussen pieken en dalen. Andere mogelijkheden om de modulatie diepte te reduceren zijn o.a. compressie en het toevoegen van ruis. Ruwweg kunnen we zeggen dat compressie de pieken verlaagt terwijl ruis de dalen verhoogt. Ruis brengt echter ook een verandering van de fijnstructuur en maskering van de zwakke spraakdelen teweeg. In hoofdstuk 4 wordt een aantal experimenten beschreven om de relatieve bijdrage van temporele modulaties en fijnstructuur aan de verstaanbaarheid te onderzoeken. Daartoe werden omhullende en fijnstructuur per $\frac{1}{4}$ -octaaf band afzonderlijk gewijzigd. Zonder de spraakfijnstructuur aan te tasten werd het belang van de omhullende bestudeerd door de laatste te onderwerpen aan een drietal bewerkingen: piek-clippen, dal-clippen en 1-bit codering (blokgolfbenadering). Het belang van de fijnstructuur werd geëvalueerd door gebruik te maken van een ruisfijnstructuur met onaangetaste spraakomhullende en spraakfijnstructuur met een door ruis aangetaste omhullende. Gewone spraak+ruis diende als referentie. Vergelijking van zinsverstaanbaarheidsscores in stilte bij 60 proefpersonen toonde het volgende aan: (1) de fijnstructuur speelt een veel minder belangrijke rol dan de omhullende; (2) de pieken in de omhullende zijn perceptief belangrijker dan de dalen; (3) reductie van de spraakmodulaties door het toevoegen van ruis is nadeliger voor de verstaanbaarheid dan dezelfde mate van reductie als gevolg van directe manipulatie van de omhullende (verschil in termen van de SRT is bijna 7 dB). Dit laatste resultaat is in strijd met het MTF-concept; een verstaanbaarheidsmodel op basis van de MTF zou voorspellen dat als de modulatiereductie gelijk is, het effect op de verstaanbaarheid ook gelijk is. Het blijkt dat ruis storende modulaties introduceert die de perceptie van de relevante spraakmodulaties bemoeilijken. Voor de in de experimenten

gebruikte bewerkingen geldt in het algemeen dat er geen één-op-één relatie tussen de MTF en de verstaanbaarheidsscores kan worden aangetoond.

In hoofdstuk 5 kijken we nader naar het effect van ruis op de verstaanbaarheid in termen van de relatieve bijdrage van spraakdelen boven (pieken) en onder (dalen) de ruis. In 24 $\frac{1}{4}$ -octaaf banden werd de verdeling van spraak en ruis over de zinnen gemanipuleerd door spraak in de pieken en ofwel ruis alleen ofwel spraak+ruis in de dalen aan te bieden. De zinsverstaanbaarheid werd voor verschillende spraak-ruisverhoudingen gemeten bij 12 luisteraars. De resultaten geven aan dat, vergeleken met gewone spraak+ruis stimuli (data uit hoofdstuk 4), het weghalen van ruis uit de pieken geen effect heeft op de verstaanbaarheid. Het weghalen van het spraaksignaal uit door ruis gedomineerde dalen geeft echter een verhoging van de SRT van 2 dB. Het blijkt dus dat spraakdelen onder het ruisniveau voor een deel bijdragen aan de verstaanbaarheid.

Alle experimentele resultaten uit hoofdstukken 2-4 zijn besproken in termen van (een reductie van) temporele modulaties. Maar manipulaties op een reeks smalbandige temporele omhullenden zorgen ook voor zekere spectrale veranderingen. In hoofdstuk 6 wordt onderzocht in hoeverre spectrale contrasten worden aangetast door de reductie van temporele modulaties. Als tegenhanger van de MTF voor het meten van de overdracht van temporele modulaties wordt een spectrale modulatie-overdrachtsfunctie (SMTF) beschreven om te bepalen in hoeverre modulaties in de spectrale omhullende (gemeten in perioden per octaaf) behouden blijven. Uniforme reductie (d.w.z. onafhankelijk van de modulatiefrequentie) van de temporele modulaties in $\frac{1}{4}$ -octaaf banden blijkt een vrijwel gelijke mate van spectrale reductie te geven. SMTF-metingen na filteren van de temporele omhullende (hoofdstukken 2 en 3) werden uitgevoerd voor een aantal laag- en hoogdoorlaat-afsnijfrequenties (1-32 Hz). Wanneer we het effect van deze temporele bewerkingen uitdrukken in termen van spectrale contrastreductie, is het interessant te onderzoeken of spectrale reductie op zich een verklaring is voor het gemeten verlies in verstaanbaarheid. Daarom werden zinnen bewerkt met een analyse-resynthesesysteem bestaande uit Fourier transformaties op een reeks korte segmenten, 10-90% reductie van de contrasten in de spectrale omhullende, inverse transformatie en overlappende optelling van de bewerkte segmenten om weer een continu signaal te krijgen. De SRT in ruis voor de op die manier spectraal gereduceerde zinnen werd gemeten bij 10 luisteraars. Vergelijking van deze SRTs met die uit hoofdstuk 2 en 3 geeft aan dat spectrale reductie hetzelfde effect op de verstaanbaarheid heeft als hoogdoorlaat filteren van de temporele omhullende. Dit geldt echter niet voor laagdoorlaat filteren, wat betekent dat spectrale reductie in dat geval slechts een bij-effect is.

References

- Bendat, J.S. and Piersol, A.G. (1980). *Engineering Applications of Correlation and Spectral Analysis* (Wiley-Interscience, New York).
- Behrens, S., and Blumstein, S.E. (1988). "On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants," *J. Acoust. Soc. Am.* **84**, 861-867.
- Bosman, A.J. (1989). "Speech perception by the hearing impaired," Ph.D. dissertation, University of Utrecht.
- Carter, G.C., Knapp, C.H., and Nuttall, A.H. (1973). "Estimation of the magnitude-squared coherence function via overlapped fast fourier transform processing," *IEEE Trans. Audio Electroacoust.* **21**, 337-344.
- van Dijkhuizen, J.N., Festen, J.M., and Plomp, R. (1989). "The effect of varying the amplitude-frequency response on the masked speech-reception threshold of sentences for hearing-impaired listeners," *J. Acoust. Soc. Am.* **86**, 621-628.
- Drullman, R., Festen, J.M., and Plomp, R. (1994a). "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.* **95**, 1053-1064.
- Drullman, R., Festen, J.M., and Plomp, R. (1994b). "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.* **95**, 2670-2680.
- Drullman, R. (1994c). "Temporal envelope and fine structure cues for speech intelligibility," accepted for publication in *J. Acoust. Soc. Am.*
- Drullman, R. "Speech intelligibility in noise: relative contribution of speech elements below and above the noise level", submitted as a Letter to the Editor in *J. Acoust. Soc. Am.*
- Drullman, R., Festen, J.M., and Houtgast, T. "Effect of temporal modulation reduction on spectral contrasts in speech," submitted for publication in *J. Acoust. Soc. Am.*
- Duquesnoy, A.J., and Plomp, R. (1980). "Effect of reverberation and noise on the intelligibility of sentences in cases of presbycusis," *J. Acoust. Soc. Am.* **68**, 537-544.
- Festen, J.M., and Plomp, R. (1981). "Relations between auditory functions in normal hearing," *J. Acoust. Soc. Am.* **70**, 356-369.
- Festen, J.M., van Dijkhuizen, J.N., and Plomp, R. (1990). "Considerations on adaptive gain and frequency response in hearing aids," *Acta Otolaryngol. Suppl.* **469**, 196-201.