Nederlandse samenvatting

Het concept van de signaal/ruis verhouding in het modulatie domein
Het voorspellen van de verstaanbaarheid van bewerkte spraak

De constatering dat de meeste ruisonderdrukkings-algoritmen niet in staat zijn het om verstaan van spraak in achtergrondruis te verbeteren, vormde de basis voor dit proefschrift. Het doel was om te onderzoeken wat de meest cruciale – signaaltechnische – factor is bij het verstaan van spraak in achtergrondruis. Een dergelijke factor is per definitie onafhankelijk van het type signaal bewerking dat wordt toegepast op het spraak-plus-ruis signaal. Recente publicaties laten zien dat dit niet het geval is voor gangbare maten zoals signaal/ruis verhouding (S/N ratio) en de speech transmission index (STI). Het is bekend dat signaal bewerking zowel de S/N ratio als de STI kunnen verbeteren, terwijl de verstaanbaarheid toch gelijk blijft. Zolang deze paradox niet is opgelost, blijft de theorie rond het spraakverstaan in ruis onvolledig, en blijft het tevens onduidelijk waar ruisonderdrukkings-algoritmen zich precies op moeten richten in termen van signaal reconstructie. Het begrijpen van deze tegenstelling is de drijvende kracht geweest achter dit promotieonderzoek, waarin de fysische en psychofysische effecten van ruis op spraak in detail zijn onderzocht. Het globale idee is dat ruis op een aantal manieren de golfvorm van het spraaksignaal verandert. Elke afzonderlijke verandering heeft een specifiek (negatief) effect op de verstaanbaarheid; de ene verandering meer dan de ander. Door het meest dominante effect te identificeren, kan de verstaanbaarheid van spraak in ruis in het algemeen wellicht worden verbeterd door juist dit effect te neutraliseren.

Na een algemene introductie in Hoofdstuk I wordt in Hoofdstuk II een speciaal type signaal analyse geïntroduceerd waarmee de verschillende veranderingen in het spraaksignaal (als gevolg van het toevoegen van ruis) systematisch kunnen worden geïdentificeerd en geïsoleerd. Om zo dicht mogelijk bij menselijke auditieve verwerking te blijven, is ervoor gezorgd dat de analyse – wat betreft de spectro-temporele eigenschappen – sterk lijkt op de signaalanalyse van het auditieve systeem. Het is bekend dat de spectrale resolutie van het auditieve systeem grofweg constant is op een logaritmische schaal, hetgeen ook geldt voor wavelet analyse. Derhalve is deze techniek gebruikt als analyse methode voor het vaststellen van de instantane amplitude en fase van een 1/3-oktaafband uitgangssignaal. Deze bandbreedte correspondeert grofweg met de bandbreedte van het auditieve systeem. Door wavelet analyse toe te passen op zowel het schone spraaksignaal als op het spraak-plus-ruis signaal, en de resulterende wavelet-pixels te vergelijken, konden drie typen effecten in de wavelet representatie van de spraak worden vastgesteld: twee effecten op de wavelet sterkte (intensiteit) en een effect op de wavelet fase. Met behulp van signaalbewerking konden elk van deze effecten afzonderlijk worden in- of uitgeschakeld. De resulterende acht mogelijke condities ($=2^3$) waren de basis voor luisterexperimenten (CVC woordscores). Hiermee werd aangetoond dat het systematische intensiteit-effect het grootste effect op verstaanbaarheid had. In deze conditie zakten de CVC scores van 83% (geen ruis toegevoegd) naar 63%. Het fase-effect bleek minder destructief (76%), en de stochastische verstoring van de intensiteits-omhullende van de spraak leek geen enkel effect te hebben (86%), behalve in combinatie met een tweede effect. Het toevoegen van een tweede effect leidde in het algemeen tot een verdere reductie in de verstaanbaarheid. Tot dit punt leken de resultaten in overeenstemming met modellen als het MFT-STI model (modulation transfer function – speech transmission index). In dit model wordt het effect van ruis op het spraakverstaan gekoppeld aan een reductie van essentiële spraakmodulaties, ofwel een afname van de piek/dal verhouding in de spraakomhullende als gevolg van ruis in de dalen van de spraak. Gemiddeld genomen neemt de energie in de dalen toe met een waarde gelijk aan de gemiddelde ruisintensiteit, hetgeen exact overeenkomt met het ‘eerste ruiseffect’ uit het experiment.

Het verbeteren van de verstaanbaarheid van spraak na het toevoegen van ruis lijkt dus eigenlijk een kwestie van het neutraliseren van het eerste ruiseffect. In signaal-technische termen is deze operatie identiek aan het opleggen van alle ruiseffecten minus het systematische intensiteit effect. Dit was een van de acht gemeten condities. In deze conditie nam de spraakverstaanbaarheid toe van 29% (het volledige ruiseffect) tot 74% na de operatie, wat het belang aangeeft van het neutraliseren van het eerste ruis effect.

Voor mogelijke applicaties in de praktijk wordt deze operatie echter sterk beperkt door het feit dat het originele (ruisvrije) spraak signaal nodig is. Een tweede, alternatieve operatie waarin een schatting van de gemiddelde ruisintensiteit direct werd afgetrokken van de spraak-plus-ruis mix – een basale

vorm van 'spectral subtraction' – bleek de STI waarde weliswaar te verbeteren (en dus het eerste ruiseffect succesvol te neutraliseren) terwijl de verstaanbaarheid toch gelijk bleef. Beide typen ruisonderdrukking (met en zonder voorkennis) bleken dus volstrekt verschillende effecten op het spraakverstaan te hebben.

Om dit te begrijpen werden de uitgangssignalen van beide typen bewerkingen vergeleken, waarbij duidelijke verschillen in signaalstructuur naar voren kwamen. Na de eerste operatie (waarbij de originele spraak werd gebruikt) werden de globale contouren van de spraakomhullende hersteld, en reduceerden de ruisfluctuaties binnen deze contouren. De tweede operatie ('spectral subtraction') compenseerde weliswaar de toename van de totale energie, maar veranderde niet wezenlijk de distributie van de ruisfluctuaties. Dus, de sterkte van de ruismodulaties werd fors gereduceerd in het eerste geval, maar bleef in essentie gelijk in het tweede geval. Dit suggereert dat spraakverstaanbaarheid wellicht niet alleen afhankelijk is van de sterkte van *spraak*-modulaties, maar ook van *ruis*-modulaties. Aldus ontstaat het beeld dat spraakverstaan in achtergrondruis wellicht afhankelijk is van de *verhouding* tussen sterkte van spraakmodulaties en ruismodulaties.

Om dit idee te verifiëren werd spraak-plus-ruis onderworpen aan verscheidene typen signaal bewerking (spectral subtraction, deterministische modulatie reductie, amplitude compressie en amplitude expansie), met als doel om de relatieve sterkte van de spraakmodulaties en de ruismodulaties in het signaal ("modulatie ratio") te vergelijken met metingen van verstaanbaarheid. Aangezien de modulatie ratio niet direct uit het uitgangssignaal zelf kan worden bepaald, werd gebruik gemaakt van een speciaal testsignaal dat, na het toevoegen van ruis, werd onderworpen aan verschillende typen signaalbewerking. Voor elk type bewerking werden modulatie ratio's bepaald en vergeleken met verstaanbaarheidscores van spraak-plus-ruis signalen die op dezelfde manier waren bewerkt (beschreven in Hoofdstuk III). De resultaten van deze studie leken de relatie tussen modulatie ratio en verstaanbaarheid te bevestigen. Aldus werd een nieuw concept geïntroduceerd: de signaal/ruis verhouding in het modulatiedomein, ofwel $(S/N)_{mod}$.

Het verbeteren van de verstaanbaarheid van spraak in ruis leek nu in essentie een kwestie van het vergroten van de $(S/N)_{mod}$, ofwel het vergroten van de relatieve sterkte van spraak- en ruismodulaties. Om dit concept verder te onderzoeken, werd de $(S/N)_{mod}$ van spraak-plus-ruis kunstmatig gevarieerd met signaalbewerking, en vervolgens vergeleken met verstaanbaarheidscores (Hoofdstuk IV). Dit keer werd het testsignaal (nodig om de $(S/N)_{mod}$ te schatten) zodanig aangepast dat de distributie van intensiteiten sterk overeenkwam met die van werkelijke spraak, zodat de $(S/N)_{mod}$ kon worden omgezet naar een te verwachten verstaanbaarheiddrempel ofwel speech reception threshold (SRT). In het eerste deel van dit hoofdstuk werd bij het opleggen van variaties in de $(S/N)_{mod}$ gebruik gemaakt van voorkennis over het originele (ruisvrije) signaal. Aangezien deze informatie enkel beschikbaar is in het laboratorium, werd tevens een alternatieve aanpak getest zonder voorkennis. De resultaten gaven aan dat de opgelegde $(S/N)_{mod}$-variaties in het algemeen werden gevolgd door variaties in het spraakverstaan (correlatiecoëfficiënt r=0.8). Niet alleen ondersteunden de resultaten het belang van het nieuwe concept, het duidde mogelijk ook op een praktische toepassing van de gebruikte signaalbewerking methode zelf, gegeven de gemiddelde +2 dB verbetering in verstaanbaarheid (zonder voorkennis) voor slechthorenden. Voor het toepassen van een dergelijk systeem in de praktijk is het nodig om te weten of verbeteringen ook worden gevonden voor meer realistische ruistypen zoals achtergrondgebabbel. Om een indruk te krijgen werd een luisterexperiment uitgevoerd (Hoofdstuk V), met realistisch achtergrondrumoer (gebabbel) opgenomen tijdens een receptie met ongeveer vijftig personen. Negen normaalhorende- en zestien slechthorende proefpersonen namen deel aan het experiment dat in opzet gelijk was aan het hiervoor beschreven experiment. De resultaten gaven aan dat de operatie nog steeds een verbetering in het spraakverstaan gaf, zowel voor de normaalhorende proefpersonen (gemiddeld +0.6 dB), als voor de slechthorende proefpersonen (gemiddeld +1.6 dB). Anderzijds volgde uit de resultaten ook dat verstaanbaarheidvoorspellingen gebaseerd op de $(S/N)_{mod}$ onbetrouwbaar zijn voor maskeerders die bestaan uit achtergrondrumoer geproduceerd wordt door slechts enkele sprekers, aangezien het concept geen rekening houdt met 'release of masking' effecten die bekend zijn bij dit type maskeerder. Deze principiële restrictie op de toepassingmogelijkheden van het $(S/N)_{mod}$ concept vereist verder onderzoek.