

Nederlandse samenvatting

In dagelijkse communicatie kunnen mensen snel schakelen tussen praten en luisteren. Spraakproductie en -perceptie wisselen elkaar in hoog tempo af. Spraakklanken verschillen in samenstelling van klinkers en medeklinkers maar bevatten ook spreker-specifieke kenmerken, zoals het geslacht en de emotie van de spreker. Ondanks variaties tussen sprekers kunnen mensen over het algemeen zonder veel moeite woorden uit spraak oppikken. Daarnaast kunnen wij ook stemmen goed herkennen, ondanks dat een persoon verschillende woorden uitspreekt. Dat doen wij onbewust door middel van het plaatsen van geluiden in multidimensionale categorieën, bijvoorbeeld gebaseerd op fonetische informatie of spreker-specifieke kenmerken. In fysiek contact, bijvoorbeeld tijdens een gesprek in een café, maken wij ook gebruik van visuele informatie om een compleet perceptueel beeld te vormen.

Auditieve categorisatie is van belang voor de perceptie van spraakklanken. Onbewust categoriseren mensen spraakklanken op verschillende dimensies, bijvoorbeeld ofwel op basis van fonetische informatie (bijvoorbeeld categorisatie op klinker) of sprekerinformatie (bijvoorbeeld categorisatie op geslacht van de spreker). Fonetische categorisatie is voor jonge kinderen van belang bij het leren van hun moedertaal en voor volwassenen is het belangrijk bij het leren van een tweede taal. Voor sociale interacties, bijvoorbeeld op een feestje waar meerdere mensen door elkaar praten, is ook het herkennen van een persoon aan zijn of haar stem belangrijk.

Gedurende mijn promotie heb ik gewerkt bij drie onderzoeksgroepen op verschillende universiteiten. De onderzoekers waarmee ik heb samengewerkt waren gespecialiseerd in gedragsbiologie, neuropsychologie, psychofysica en/of neuro-imaging. In mijn multidisciplinaire projecten heb ik onderzocht wat de overeenkomsten en verschillen zijn tussen taalgerelateerde cognitieve eigenschappen bij mensen en zangvogels. Daarnaast heb ik onderzocht in hoeverre een verschil tussen wat wij horen (het gesproken woord of de stemklank) en wat wij zien (informatie die af te lezen is aan het gezicht) onze waarneming vervolgens kan veranderen. Tot slot heb ik bestudeerd wat er in de hersenen gebeurt wanneer mensen naar spraakklanken luisteren die verschillen in klinker en geslacht van de spreker.

In het eerste deel van dit proefschrift presenteer ik twee studies over de categorisatie van spraak(achtige) geluiden en de onderliggende mechanismen bij een zangvogelsoort (de zebra-vink). Alle diersoorten hebben hun eigen manier van communicatie en het is gemakkelijk verschillen aan te wijzen tussen bijvoorbeeld de menselijke taal en de communicatie tussen zangvogels. Er zijn echter ook duidelijke overeenkomsten tussen deze vormen van communicatie. Daarom is het interessant om te onderzoeken in hoeverre dieren de vaardigheden, die van belang zijn bij menselijke taal, kunnen beheersen of leren. Met behulp van gedragsonderzoek bij dieren kunnen wij reconstrueren welke cognitieve mechanismen gemeenschappelijk zijn en mogelijk gedeeld zijn met onze gemeenschappelijke

Samenvatting

voorouders. In hoeverre kunnen dieren ook spraakklanken onderscheiden en op welke manier doen zij dat?

Categorisatie is een algemeen perceptueel fenomeen dat essentieel is voor alle dieren om de verschillende perceptuele input (waarnemingen) te begrijpen. Categorisatie gebaseerd op fonetische informatie of spreker-specifieke kenmerken is belangrijk voor menselijke taal. Maar hoe specifiek is dit voor menselijke taal? Eerdere studies hebben aangetoond dat verschillende (vogel)soorten, waaronder zebra's, kunnen leren om spraakklanken te onderscheiden op basis van klinkerverschillen (bijvoorbeeld 'wet' versus 'wit'). Daarna konden zij ook deze spraakklanken van een andere spreker onderscheiden. Naar aanleiding van deze resultaten vroegen wij ons af hoe de zebra's deze indeling kunnen maken en of zij ook in staat zijn om te categoriseren op basis van de andere relevante dimensie, namelijk het geslacht van de spreker.

In HOOFDSTUK 2 beschrijf ik een vergelijkend onderzoek naar zebra's en mensen. Wij gebruikten vier categorisatietaken die wij ééndimensionaal of tweedimensionaal noemden. De vogels en proefpersonen kregen willekeurig één van de vier taken toegewezen. Alle vier de taken begonnen met een training. In de training van de eendimensionale taken moesten de proefpersonen of vogels leren om vier spraakklanken op basis van één dimensie in twee categorieën te delen, ofwel op klinker (*wet* versus *wit*) of op stemklank (man versus vrouw). Wanneer het antwoord juist was, kregen vogels wat te eten als beloning en proefpersonen kregen op het computerscherm te zien dat zij goed hadden geantwoord. In de tweedimensionale taak kregen proefpersonen of vogels een beloning wanneer zij de spraakgeluiden categoriseerden op basis van beide dimensies (*wet* uitgesproken door de man en *wit* uitgesproken door de vrouw versus *wit* uitgesproken door de man en *wet* uitgesproken door de vrouw). Deze taak kan worden opgelost door gebruik te maken van ofwel een combinatieregel of door de twee dimensies te integreren (een categorie die wordt gedomineerd door *wet* uitgesproken door de man versus een categorie die wordt gedomineerd door *wit* uitgesproken door de vrouw). Daarna volgde een testfase waarin de proefpersonen of vogels meer ambiguë of juist heldere variaties van de geluiden moesten indelen. De vogels konden nog steeds beloningen verdienen door de trainingsgeluiden goed te categoriseren. Voor het juist categoriseren van de testgeluiden kregen zij geen beloning.

Hoewel zebra's veel langer deden over de training, waren zowel mensen als zebra's sneller in het leren van de ééndimensionale taken dan de tweedimensionale taken. In de testfase werd het verschil tussen soorten groter. Vogels waren goed in het onderscheiden van de trainingsgeluiden, wat suggereert dat zij goed zijn in voorbeelden onthouden. Mensen konden vooral de nieuwe, heldere testgeluiden, die het meest op echte spraak lijken, goed onderscheiden. Eén van de verklaringen voor dit resultaat is de levenslange ervaring van mensen met het categoriseren op basis van klinker en geslacht van de spreker. Bovendien hebben mensen vaak de neiging om een op regels gebaseerde aanpak te hanteren, wat ook blijkt uit de moeite die mensen hebben met de taak waarbij geen simpele regel kan worden toegepast.

Op welke manier kunnen de zebra's spraakklanken onderscheiden? En lijkt dat op de manier waarop mensen dat doen? Onbewust letten mensen op verschillende aspecten van geluid, geluidsparemeters, wanneer wij spraakklanken categoriseren. Voor het categoriseren van stemmen, letten wij vooral op de toonhoogte (pitch) van de stem. Klinkers onderscheiden wij vooral op basis van timbre, wat vooral wordt bepaald door de versterkte frequentiebanden, de formanten, in het spraakgeluid. Hoewel vogels niet dagelijks te maken hebben met variabiliteit in menselijke spraak, hebben zij wel dagelijks te maken met variatie in de liedjes van hun soortgenoten die ook variëren in toonhoogte en harmonische kenmerken.

Nadat wij in de studie in HOOFDSTUK 2 hadden geleerd dat vogels in staat zijn menselijke spraakklanken te onderscheiden op basis van klinker en geslacht van de spreker, besloten wij om in de studie in HOOFDSTUK 3 verder te onderzoeken hoe zebra's verschillende geluidsparemeters gebruiken bij het categoriseren van spraakachtige geluiden. Tijdens de training moesten de vogels leren om kunstmatige spraakklanken, die nog het meest leken op klinkers, in twee categorieën te sorteren. De helft van de onderzochte vogels moest categoriseren op basis van fundamentele frequentie, f_0 (lage versus hoge 'toonhoogte'); de andere helft van de onderzochte vogels moest dat doen op basis van spectrum (laag versterkte frequentieband, 'formant', versus hoog versterkte frequentieband). Nadat de vogels de training succesvol hadden afgerond, kregen zij tijdens de testfase geluiden die alleen informatie over de toonhoogte bevatten, zonder een versterkte frequentieband. Daarnaast moesten de vogels in deze testfase geluiden categoriseren die alleen informatie bevatten over het globale spectrum maar geen informatie over de fijne spectrale details en toonhoogte. Deze testgeluiden waren gemaakt door de trainingsgeluiden met een vocoder, een computer algoritme, te bewerken.

De zebra's leerden tijdens de training onderscheid te maken tussen kunstmatige klinkerachtige geluiden op basis van verschillen in toonhoogte of op basis van verschillen in het globale spectrum. Het verschil in training had het verwachte effect tijdens de testfase. De geluidsparemetere waar de zebra's op focussen is afhankelijk van de training. Na een training op toonhoogte konden de vogels generaliseren naar geluiden zonder een versterkte frequentieband maar zij maakten geen onderscheid tussen de testgeluiden met informatie over het globale spectrum. Nadat de vogels waren getraind op spectrum konden de vogels niet generaliseren naar de testgeluiden zonder versterkte frequentieband maar zij maakten wel onderscheid tussen de testgeluiden met informatie over het globale spectrum en deze responsen kwamen overeen met hun categorisatie van de trainingsgeluiden.

Deze zangvogels zijn dus erg flexibel in hun perceptuele strategie en zij gebruiken de meest relevante paremeters om een auditieve categorisatietaaK op te lossen. Onze bevindingen zetten het traditionele idee, namelijk dat zoogdieren en zangvogels fundamenteel verschillen wat betreft de auditieve verwerking van toonhoogte (pitch), in een ander daglicht. Mogelijk letten zangvogels in eerste instantie wel op de precieze hoogte van de pitch maar zij kunnen

Samenvatting

ook de toonhoogte in verhouding tot andere geluiden waarnemen, net zoals veel zoogdieren (inclusief mensen, ratten en fretten) doen.

In het tweede deel van dit proefschrift ben ik verder ingegaan op cognitieve en neurale mechanismen die ten grondslag liggen aan audio(visuele) verwerking van spraak bij mensen. In dagelijkse communicatie halen mensen niet alleen veel informatie uit de gesproken taal maar ook uit visuele informatie van de spreker. In minder dan een halve seconde integreren wij informatie van de stem die wij horen met het gezicht dat wij zien. In natuurlijke omstandigheden zijn gezicht en stem over het algemeen duidelijk met elkaar in overeenstemming maar soms zijn er kleine verschillen tussen wat wij horen en zien, meestal omdat één van de twee signalen onduidelijk of dubbelzinnig is. In hoeverre kunnen wij in situaties waarin auditieve en visuele informatie niet overeenkomen toch een compleet beeld vormen? Kunnen wij onze categorieën aanpassen met betrekking tot klinker en geslacht van de spreker?

Tot slot hebben wij bestudeerd wat er in de hersenen gebeurt wanneer mensen naar spraakklanken luisteren die verschillen in klinker en geslacht van de spreker. Er is al bekend welke hersengebieden betrokken zijn bij klinker en spreker verwerking. Het is nog onduidelijk of deze hersengebieden in de auditieve cortex alleen bijdragen aan de verwerking van akoestische eigenschappen van het geluid of ook al een functie vervullen bij het categoriseren van spraak op basis van klinkers en sprekers. In hoeverre kunnen wij de activatie patronen in het brein verklaren met behulp van een model op basis van akoestische eigenschappen, een ander model op basis van categorische eigenschappen of een derde model op basis van gedragsresponsen?

In onze studie, beschreven in HOOFDSTUK 4, onderzochten wij in welke mate mensen hun categorieën kunnen aanpassen op basis van eerdere blootstelling aan ofwel auditieve en visuele informatie die bij elkaar passen (bijvoorbeeld een mannelijk gezicht in combinatie met een mannelijke stem) of visuele informatie in combinatie met een ambigue geluid (bijvoorbeeld een mannelijk gezicht in combinatie met een androgyne stem). In het eerste geval, wanneer geluid en beeld tijdens blootstelling overeenkomen en de persoon zijn of haar categorisatie van het volgende geluid daarop aanpast, spreken wij van selectieve adaptatie. In het tweede geval, wanneer beeld en geluid tijdens blootstelling niet overeenkomen en de persoon zijn of haar categorisatie van het volgende geluid daarop aanpast, spreken wij van recalibratie.

Deelnemers kregen een aantal korte video's te zien waarin een mannelijke of vrouwelijke spreker 'beek' of 'beuk' zei. Het geluid bij de video was ofwel helder en overeenkomstig met de visuele informatie van het gezicht (selectieve adaptatie) of ambigue (recalibratie) met betrekking tot de klinker of het geslacht van de spreker. De geluiden die ambigue klonken voor klinker waren een mix van de woorden *beek* en *beuk*. Voor geslacht van de spreker klonk dat ambigue, androgyne geluid als een mix van een mannelijke en vrouwelijke stem. Daarna

kregen proefpersonen ambigue testgeluiden te horen die zij moesten categoriseren op basis van klinker of geslacht van de spreker.

In lijn der verwachtingen, gaven de proefpersonen na blootstelling aan een video van *beek* (respectievelijk *beuk*) gecombineerd met een ambigue klinkend woord vaker aan dat zij het ambigue testgeluid vonden klinken als *beek* (respectievelijk *beuk*). Dit resultaat is een aanwijzing dat mensen hun categorieën voor klinkers kunnen recalibreren. In overeenstemming met eerder onderzoek reageerden proefpersonen vaker met *beuk* (respectievelijk *beek*) wanneer zij waren blootgesteld aan een video van een *beek* (respectievelijk *beuk*) gecombineerd met het bijpassende gesproken woord (selectieve adaptatie voor klinker).

Zoals wij hadden verwacht, gaven de proefpersonen na blootstelling aan een video van een man (respectievelijk vrouw) in combinatie met een androgyne stem vaker aan dat zij het ambigue testgeluid mannelijk (respectievelijk vrouwelijk) vonden klinken. Dit resultaat is een aanwijzing dat mensen hun categorieën recalibreren voor het geslacht van de spreker. In overeenstemming met eerder onderzoek, gaven de proefpersonen meer 'vrouw' responsen (respectievelijk 'man' responsen) na blootstelling aan video van een man (respectievelijk vrouw) wanneer deze video gepaard was met de passende stem van deze spreker (selectieve adaptatie voor het geslacht van spreker).

Hieruit volgt dat wij recalibratie en selectieve adaptatie voor zowel klinker als geslacht van de spreker hebben aangetoond. Mensen kunnen hun categoriegrenzen flexibel aanpassen om een samenhangend beeld van de auditieve en visuele input te maken.

Nadat wij in de studie in HOOFDSTUK 4 hadden geleerd dat mensen ook ambigue geluiden van *beek/beuk* van een mannelijke en vrouwelijke spreker kunnen categoriseren, besloten wij deze geluiden te gebruiken in de fMRI-studie in HOOFDSTUK 5 waarin wij de activatiepatronen in het brein onderzochten wanneer mensen naar spraakklanken luisterden die varieerden in klinker en spreker. Wij gebruikten in deze studie alleen de klinkers (*ee* en *eu*) van de vier geluidsopnames van '*beek*' en '*beuk*' uitgesproken door een mannelijke en vrouwelijke spreker uit de studie in HOOFDSTUK 4. Op basis daarvan maakten wij nog een aantal extra ambigue en heldere variaties. Na een oefensessie in de scanner, vroegen wij aan deelnemers of zij tijdens de MRI-scan een detectietaak wildendoen waarbij zij moesten aangeven wanneer een afgespeeld geluid langer duurde dan de andere geluiden. Na de MRI-scan deden de deelnemers twee gedragstaken op de computer waarbij zij dezelfde geluiden op basis van klinker en spreker moesten categoriseren. De met fMRI gemeten BOLD-responsen van de deelnemers werden op groepsniveau geanalyseerd met behulp van drie modellen: een model op basis van de continue akoestische eigenschappen, een ander model op basis van categorische eigenschappen (*ee* of *eu* categorie en man of vrouw categorie) of een derde model op basis van gedragsresponsen van deelnemers tijdens de categorisatietaken na de MRI-scan.

Samenvatting

In de klinker en spreker categorisatietaken werden de gedragsresponsen zowel door (de ambiguïteit van) klinkers als die van de sprekers van het spraakgeluid significant beïnvloed. Alle drie de modellen toonden aan dat de BOLD-activering in bilaterale auditieve cortices in vergelijkbare mate verklaard kon worden door voorspellers van klinkers en sprekers. Dit suggereert dat sensorische en categorische processen die relevant zijn voor de perceptie van klinkers en sprekers nauw met elkaar verbonden zijn. De verschillen tussen hersengebieden die betrokken zijn bij akoestische verwerking of categorisatie lijken zo subtiel te zijn dat er op groepniveau lastig onderscheid te maken is.

De studies in dit proefschrift laten zien dat multidisciplinaire studies ons meer inzicht kunnen geven in de cognitieve mechanismen, audiovisuele mechanismen en activatiepatronen in het brein die bijdragen aan de categorisatie van klinkers en sprekers. Door (zang) vogelsoorten te vergelijken met mensen, kunnen wij taalspecifieke mechanismen ontrafelen die ten grondslag liggen aan spraakperceptie en categorisatie. Door het onderzoeken van verschillende audiovisuele processen bij dezelfde mensen, kunnen wij de domein algemene versus domeinspecifieke aspecten van audiovisuele recalibratie verder onderzoeken. Ten slotte zou men met behulp van individuele analyses de complexiteit kunnen ontrafelen van de fijnmazige patronen die betrokken zijn bij sensorische en categorische processen die verband houden met klinker- en sprekerperceptie. Door het bestuderen van diergedrag en het combineren van psychofysica en neuro-imaging kunnen wij meer ontdekken over gedeelde en unieke cognitieve, audiovisuele en neurale mechanismen die ten grondslag liggen aan spraak categorisatie.