



Universiteit
Leiden
The Netherlands

Expression and recognition of emotion in native and foreign speech : the case of Mandarin and Dutch

Zhu, Y.

Citation

Zhu, Y. (2013, December 12). *Expression and recognition of emotion in native and foreign speech : the case of Mandarin and Dutch*. LOT dissertation series. Retrieved from <https://hdl.handle.net/1887/22850>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/22850>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/22850> holds various files of this Leiden University dissertation.

Author: Zhu, Yinyin

Title: Expression and recognition of emotion in native and foreign speech : the case of Mandarin and Dutch

Issue Date: 2013-12-12

Samenvatting

Vanaf het moment dat Darwin zijn boek *De uitdrukking van Emoties bij Mens en Dier* [*The Expression of the Emotions in Man and Animals*] uitgaf in 1872, heeft een niet aflatende stroom publicaties over de productie en perceptie van emoties het licht gezien. Eerder onderzoek is vooral uitgevoerd onder de vlag van psychologie, fysiologie, biologie; maar verlegde allengs zijn grenzen en omvatte daarna mede de sociologie, taalwetenschap, pathologie, informatica, neurowetenschap, musicologie en tweedetaalverwerving. Daarenboven is er altijd al onderzoek gedaan naar de waarneming of uitdrukking van emoties vanuit over verschillende talen (cross-linguïstisch) en culturen (cross-cultureel) heen. Volgens sommige onderzoekers is de waarneming van emoties universeel. Andere onderzoekers menen dat de waarneming deels universeel verloopt en deels taal-of-cultuurspecifiek is. Eerder onderzoek laat zien dat emoties over het algemeen beter herkend worden als deze worden uitgedrukt door een spreker van dezelfde taal en afkomstig uit dezelfde cultuur als de luisteraar; dit wordt het in-groepvoordeel genoemd. Deze onderzoeken hebben ook uitgewezen dat automatische herkenning van door mensen geproduceerde emoties akoestische eigenschappen kan aanwijzen die menselijke luisteraars gebruiken om emoties te herkennen en als spreker mee uit te drukken. Ondanks de veelheid aan onderzoek naar de productie en perceptie van emoties, krijgen we uit het eerdere onderzoek geen goed beeld van hoe goed luisteraars van een niet-toontaal emoties kunnen waarnemen die worden uitgedrukt in een toontaal, vooral als die uitdrukking plaats heeft via het auditieve kanaal. Evenmin geeft het eerder onderzoek aan hoe succesvol niet-moedertaalsprekers van een taal vocale emoties kunnen produceren in een tweede of vreemde taal (T2), in verhouding tot wat moedertaalsprekers kunnen, vooral niet wanneer de vreemde taal (T2) een toontaal is maar de moedertaal (T1) van de spreker niet. We weten ook niet of een spreker vocale emoties in zijn T2 even goed kan uitdrukken als in zijn moedertaal (T1).

Daarom was het eerste doel van dit proefschrift om experimenteel te onderzoeken hoe goed Chinese moedertaalsprekers en Nederlandse niet-moedertaalsprekers van het Mandarijn (Standaard Chinees) vocaal uitgedrukte emoties kunnen waarnemen in een toontaal, in casu het Mandarijn. De niet-moedertaalluisteraars waren hetzij naïeve luisteraars hetzij gevorderde Nederlandse leerders van het Mandarijn (met dezelfde moedertaalachtergrond als de naïeve luisteraars). In de tweede plaats heb ik onderzocht of Nederlandse leerders van het Mandarijn even goed in staat zijn vocale emoties te produceren in de vreemde taal als in hun eigen niet-toontaal, d.w.z. het Nederlands; tegelijkertijd heb ik nagegaan hoe goed naïeve moedertaalluisteraars en gevorderde leerders van het Mandarijn vocale emoties kunnen waarnemen die worden uitgedrukt door L2-sprekers van de toontaal. Naderhand heb ik een akoestische analyse uitgevoerd om de vocale correlaten op te sporen die sprekers en luisteraars gebruiken bij resp. de productie en perceptie van de vocale emoties. Als laatste heb ik onderzocht of het in-groepvoordeel dat door andere onderzoekers gerapporteerd is (waarbij de claim was

luisteraars in het algemeen emoties succesvoller herkennen in hun eigen taal dan in een onbekende taal), universeel (d.w.z. taal- en cultuuronafhankelijk) is.

Vanuit theoretisch gezichtspunt probeert dit proefschrift een functionele hypothese te toetsen die claimt dat de prosodische ruimte die talen kunnen gebruiken eindig is. Als een taal bv. duur gebruikt om een segmenteel contrast te markeren tussen korte en lange klinkers, dan zal de duurparameter geen rol kunnen spelen (of hooguit een geringe rol) bij de markering van klemtoon – welk verschijnsel in andere talen sterk leunt op duuraanwijzingen (Berinstein 1979, Potisuk et al. 1999, Remijsen 2002a, b). Volgens dezelfde redenering zal een taal zoals het Mandarijn, die toonhoogteverschillen gebruikt om verschillen tussen woordvormen te signaleren (zgn. lexicale toon), minder ruimte over hebben om met behulp van toonhoogte paralinguïstische contrasten te markeren, bv. om emoties uit te drukken. Dientengevolge voorspel ik dat moedertaal-luisteraars van het Mandarijn slechts beperkt zijn blootgesteld aan duidelijke voorbeelden van prosodisch uitgedrukte emoties. Meer algemeen voorspel ik dat luisteraars met een toontaalachtergrond minder gespitst zullen zijn op (en minder ervaren zullen zijn in) het decoderen van paralinguïstisch gebruik van prosodie dan luisteraars met een niet-toontaal als moedertaal.

In het inleidend **Hoofdstuk 1** vat ik samen wat bekend is uit eerder onderzoek en identificeer ik enkele onopgeloste vragen. Ook verschaft ik in dit hoofdstuk achtergrondinformatie over het verschil tussen toontalen en niet-toontalen, specifiek over het Mandarijn Chinees en het Nederlands, over toon en (emotionele) prosodie en over de akoestische eigenschappen van prosodie. Ik geef vervolgens een puntsgewijs overzicht van mijn onderzoeksvragen en leid specifieke voorspellingen af op grond van de functionele zienswijze die ik hierboven heb genoemd. Ook geef ik mijn beweegredenen voor de keuze van zes emotionele prosodieën die ik in deze studie gebruik ('neutraal', 'blij', 'boos', 'verbaasd', 'bedroefd' en 'sarcastisch') en stel ik haalbare onderzoeksmethoden voor om antwoorden te vinden op mijn onderzoeksvragen. Aan het eind van dit hoofdstuk schets ik de organisatie van de rest van het proefschrift.

Hoofdstuk 2 geeft een meer gedetailleerde bespreking van eerder onderzoek. Er is veel onderzoek gedaan naar de waarneming van emotionele prosodie binnen een en dezelfde cultuurgroep. Onderzoek naar de herkenning van vocale emotie over culturen heen (zgn. cross-culturele herkenning) is slechts mondjesmaat verricht en er is nog minder vergelijkend onderzoek gedaan naar de productie van emotie. Er lijkt in het geheel geen eerder onderzoek te bestaan dat een directe vergelijking maakt van de productie van emotionele prosodie in de sprekers T1 (moedertaal) en T2 (vreemde taal). In dit laatste opzicht betreedt het productiedeel van dit proefschrift onontgonnen terrein.

Veel onderzoekers hebben beweerd dat luisteraars emotionele prosodie boven kans kunnen herkennen zowel binnen hun eigen cultuur als over culturen heen. Emoties worden over het algemeen beter herkend als sprekers en luisteraars tot dezelfde culturele groep behoren. Herkenning wordt minder trefzeker naar mate de culturele afstand tussen spreker- en luisteraargroep groter is. Eerder onderzoek heeft bovendien laten zien dat succesvolle communicatie van vocale emotie zowel van de spreker als van de luisteraar afhangt, al lijkt de rol van de luisteraar van groter belang dan die van de

spreker. Sommige onderzoekers geven aan dat de productie van emotionele prosodie tot op zekere hoogte universeel (algemeen menselijk) maar dat de productie van sommige emoties eerder cultuurspecifiek is.

In dit hoofdstuk, geef ik ook informatie over de methodologie die ik gebruik. Het proefschrift omvat drie beoordelingsstudies. De eerste beoordelingsstudie is een luisterexperiment (Exp. 1) waarin moedertaalluisteraars van het Chinees, naïeve Nederlandse luisteraars en gevorderde Nederlandse studenten Chinees aan de hand van de prosodie (d.w.z. melodie en temporele organisatie) de boven genoemde zes Chinese emoties voortgebracht door moedertaalsprekers van het Chinees moesten beluisteren en herkennen. Dit experiment beoogt antwoord te geven op onderzoeksvraag (i) *Hoe goed kunnen Chinese moedertaalluisteraars, naïeve Nederlandse luisteraars en gevorderde Nederlandse leerders van het Chinees Chinese emotionele prosodische vormen herkennen wanneer die vocaal neergezet zijn door moedertaalsprekers van het Chinees. En wat zijn de verwarringspatronen bij elk van deze drie luisteraargroepen?* Het tweede beoordelingsexperiment viel uiteen in twee perceptieproeven: in de eerste luisterden dezelfde drie luisteraargroepen naar dezelfde zes Chinese emotionele prosodische vormen maar die waren nu geproduceerd door gevorderde Nederlandse studenten Chinees (Exp. 2A). Dit experiment was opgezet om antwoord te vinden op onderzoeksvraag (ii) *Hoe goed kunnen Chinese moedertaalluisteraars, Nederlandse naïeve luisteraars en gevorderde Nederlandse leerders van het Chinees Chinese emotionele prosodische vormen herkennen die vocaal zijn neergezet door Nederlandse studenten Chinees. En wat zijn de verwarringspatronen bij de drie luisteraargroepen?* In de tweede perceptieproef (Exp. 2B) luisterden Nederlandse moedertaalluisteraars naar dezelfde zes emotionele prosodische vormen die nu waren geproduceerd in hun eigen taal (Nederlands) door dezelfde Nederlandse studenten Chinees. Met deze proef kon ik nagaan hoe doeltreffend dezelfde Nederlandse studenten Chinees de emotionele prosodische vormen in hun eigen taal kunnen produceren. De resultaten van Exp. 2A en Exp. 2B zijn met elkaar vergeleken om antwoord te vinden op onderzoeksvragen (iii) *Kunnen Nederlandse T2-sprekers van het Chinees emotionele prosodische vormen in the L2 even goed produceren als in hun T1 – Nederlands? Wat zijn de overeenkomsten en verschillen tussen deze twee productietypen?* en (iv) *Vormt de T2 een beperking bij de expressie van emotionele prosodie, vooral wanneer de T2-sprekers van de toontaal een niet-toontaal als moedertaal hebben. non-tonal language?* Onderzoeksvraag (v) *Is er steun voor de functionele hypothese die voorspelt dat luisteraars met een toontaalachtergrond minder gespitst zijn op (en minder ervaren zijn in) het decoderen van paralinguïstische onderscheidingen dan luisteraars met een niet-toontaal als achtergrond?* kan worden beantwoord aan de hand van het eerste en het tweede beoordelingsexperiment. Vervolgens is een akoestische analyse gemaakt van de stimuli die in de twee beoordelingsexperimenten zijn aangeboden. De resultaten van deze analyse beantwoordden onderzoeksvraag (vi) *Welke akoestische parameters dragen bij aan het onderscheid tussen de zes emotional prosodische vormen in het algemeen? Welke akoestische correlaten gebruiken sprekers en luisteraars als zij deze vocale emoties produceren en waarnemen in hun T1 en in hun T2? Passen de Nederlandse T2-sprekers van het Chinees transfer uit de moedertaal toe om emotionele prosodische vormen te produceren in het Chinees? In hoeverre weerspiegelt automatische herkenning de waarneming van de emotionele prosodische vormen door de menselijke luisteraars?* Het derde beoordelingsexperiment was symmetrisch opgezet. Het omvatte twee perceptieproeven waarin onervaren Chinese en Nederlandse luisteraars de zes emoties te horen kregen zoals die vocaal waren uitgedrukt in hun moedertaal en in de andere taal (Exp. 3). Dit experiment was opgezet om antwoord te krijgen op onderzoeksvraag (vii) *Is het in-groepvoordeel universeel, d.w.z. herkennen luisteraars emotionele*

prosodie altijd beter als die wordt geproduceerd in hun moedertaal dan in een tweede of geheel onbekende taal? Is de herkenning van vocale emoties cross-cultureel symmetrisch bij Chinese en Nederlandse luisteraars, m.a.w. kunnen Nederlandse en Chinese luisteraars emotionele prosodische vormen in de andere taal even trefzeker herkennen? De drie beoordelingsexperimenten tezamen geven antwoord op onderzoeksvraag (viii) Zijn productie en perceptie van emotionele prosodie universeel? Of zijn ze eerder taal- en/of cultuurspecifiek?

Hoofdstuk 3 doet verslag van het eerste beoordelingsexperiment (Exp. 1), dat tevens dient als vergelijkingsconditie voor de latere proeven. Twintig Chinese moedertaal-luisteraars, 20 naïeve Nederlandse luisteraars en 20 gevorderde Nederlandse L2 leerders van het Chinees moesten de zes Chinese emotionele prosodische vormen (neutraal, blij, boos, verbaasd, bedroefd en sarcastisch) herkennen. De resultaten laten zien dat de gevorderde Nederlandse T2 leerders van het Chinees de Chinese emotionele prosodie significant (54 % correct) beter herkenden dan de Chinese moedertaalluisteraars zelf (46 % correct) en ook beter dan Nederlandse naïeve luisteraars (46 % correct). De resultaten laten ook zien dat naïeve T2 (Nederlandse) luisteraars emoties in een onbekende taal (Mandarijn) even goed kunnen herkennen als de moedertaalluisteraar zelf. De Chinese moedertaalluisteraars gaven dus geen blijk van een in-groepvoordeel: zij identificeerden de Chinese emoties niet trefzekerder en met meer zelfvertrouwen. Voor alle drie de luisteraargroepen was ‘neutraal’ de gemakkelijkste emotie om te identificeren; ‘boos’ werd door alle luisteraargroepen even vaak correct herkend. De voorspelling die ik gedaan heb aan het begin van mijn onderzoek is hiermee bevestigd: luisteraars met een toontaalachtergrond zijn minder bedreven in het paralinguïstisch gebruik van prosodie dan luisteraars die geen toontaal als achtergrond hebben. De resultaten van dit hoofdstuk vormen de nullijn waartegen de resultaten van het volgende hoofdstuk tegen zullen worden afgezet.

Hoofdstuk 4 onderzoekt de verschillen in perceptie van zes Chinese emotionele prosodische vormen (neutraal, blij, boos, verbaasd, bedroefd en sarcastisch) die werden geproduceerd door Nederlandse T2-sprekers van het Chinees en die welke waren ingesproken door moedertaalsprekers van het Chinees (controlegroep). Dit hoofdstuk vergelijkt de resultaten van Exp. 1 en Exp. 2A. Twintig Chinese T1-luisteraars, 20 naïeve Nederlandse T2-luisteraars en 20 gevorderde Nederlandse studenten Chinees beluisterden de Chinese emotionele prosodische vormen die waren uitgedrukt door zowel T1- als T2-sprekers van het Chinees. De resultaten laten zien dat de drie luisteraargroepen de emotionele prosodische vormen die waren ingesproken door Chinese moedertaalsprekers significant beter (49% correct) herkenden dan die welke waren geproduceerd door L2-sprekers van het Chinees (39% correct). Ook blijkt dat de naïeve T2-luisteraars de emoties in de onbekende taal even goed konden herkennen (46% correct) als de Chinese moedertaalluisteraars zelf. Hoewel de gevorderde Nederlandse studenten van het Chinees het iets beter deden dan de twee overige luisteraargroepen, was er geen significant verschil tussen de drie luisteraargroepen in de identificatie van emotionele prosodie die is geproduceerd door T2-sprekers. De functionele voorspelling is hiermee opnieuw bevestigd: luisteraars met ene toontaalachtergrond zijn minder bedreven in het paralinguïstisch gebruik van prosodie dan luisteraars met een niet-toontaalachtergrond, of a non-tonal language. Er zijn dus situaties waarin het taalkundig gebruik van een akoestische eigenschap beperkingen oplegt aan het gebruik van die eigenschap ten behoeve van de communicatie van emotie.

Hoofdstuk 5 presenteert een completer beeld van hoe Nederlandse T2 sprekers van het Chinees de zes emotionele prosodische vormen in de T2 (Chinees) produceren en hoe zij dezelfde emotionele prosodische vormen in hun moedertaal (Nederlands) uitdrukken. Dit hoofdstuk rapporteert de resultaten van het eerste en het tweede beoordelingsexperiment op een geïntegreerde wijze vanuit het oogpunt van de productie. De resultaten laten zien dat emotionele prosodische vormen geproduceerd door Nederlandse T2-sprekers van het Chinees over de gehele linie minder herkenbaar zijn (39% correct door Chinese luisteraars) in hun T2 dan die welke door Chinese moedertaalsprekers werden uitgesproken (46% correct). Nederlandse T2-sprekers van het Chinese slaagden er beter in vocale emoties te produceren in hun moedertaal (57% correct geïdentificeerd door Nederlandse luisteraars). De voorspelling die ik aan het begin van dit hoofdstuk heb geformuleerd, is hiermee bevestigd: spreken in een vreemde taal beperkt de sprekers expressie van emotie. De resultaten laten ook zien dat de naïeve Nederlandse luisteraars in staat waren de emoties in de onbekende taal even goed te herkennen als de moedertaalluisteraars van het Mandarijn zelf. Bovendien lieten de naïeve Nederlandse luisteraars een in-groepvoordeel zien: zij identificeerden de emoties in hun eigen taal (Nederlands) accurater dan die in het Chinees.

Hoofdstuk 6 bevat een akoestische analyse van drie typen productie van emotionele prosodie: T1-Mandarijn, T2-Mandarijn en T1-Nederlands (de twee laatste typen waren geproduceerd door dezelfde individuen). Acht akoestische correlaten zijn onderzocht: spreektempo, gemiddelde grondfrequentie (toonhoogte of F0), de standaard deviatie van de toonhoogte (SD_F0), snelheid van F0-verandering (helling_F0), compactheid van de spectrale energieverdeling, standaard deviatie van de intensiteit (SD_int), jitter (jengel, d.w.z. de cyclus-op-cyclusvariatie van de periode van de stembandpuls, als maat voor stemvastheid) en HNR (harmonischen-ruisverhouding, als maat voor de ruizigheid van de stembandpuls). Ik heb ook een automatische herkenning van de zes emotionele prosodische vormen zoals geproduceerd door de drie sprekergruppen uitgevoerd met behulp van Lineaire Discriminant Analyse (LDA). De akoestische analyse toont aan dat de grondfrequentieparameters, d.w.z. gemiddelde F0, SD_F0 en F0-helling, van grote invloed zijn bij de productie van vocale emotie door de drie sprekergruppen. Deze bevinding komt overeen met een onderzoek van Scherer (1996), waarin eveneens is gevonden dat F0 een cruciale rol speelt bij de productie van emotionele prosodie. Jitter en standaard deviatie van de intensiteit dragen weinig bij aan het onderscheid tussen de emoties in mijn onderzoek. 'Tempo' en 'compactheid' zijn alleen onderscheidend bij moedertaalsprekers van het Mandarijn. Helling van de F0 wijst uit dat Chinese sprekers een stijgende intonatie gebruiken om verbazing uit te drukken, wat eerdere studies bevestigt die stelden dat veel toontalen verbazing uitdrukken met stijgende toonhoogte (Yip 2006). Daarenboven onderscheidt HNR duidelijk 'bedroefd' van 'neutraal' in T2-Mandarijn en in T1-Nederlands (geproduceerd door dezelfde Nederlandse individuen) maar niet in T1-Mandarijn. Samenvattend: grondfrequentie is een heel invloedrijke variabele in de productie en perceptie van vocale emotie in het algemeen. Andere parameters die in dit hoofdstuk onderzocht zijn, dragen eveneens bij aan het onderscheid tussen emotionele prosodische vormen, maar deze zijn meer emotiespecifiek en/of sprekersspecifiek.

De resultaten van de LDA geven aan dat de emotionele prosodie die door menselijke sprekers is voortgebracht, automatisch ruim boven kans herkend kan worden aan de

hand van de uitgevoerde akoestische metingen (totaal 50% correcte identificatie). Er blijkt een significante correlatie tussen de perceptieve verwarringen van emotionele categorieën bij de menselijke luisteraars en die van de automatische herkenning in het huidige onderzoek. Dit geeft aan dat de automatische herkenning tot op zekere hoogte een model is van de menselijke perceptie van deze emoties. Dit sluit echter geenszins uit dat ook nog andere akoestische parameters een bijdrage kunnen leveren aan de productie en de perceptie van emotionele prosodie in het algemeen, die wij in deze dissertatie niet heb kunnen identificeren.

De akoestische analyse wijst uit dat de Nederlandse T2-sprekers bij de productie van emotionele prosodie in hun T2 (Chinees) sommige akoestische parameters op dezelfde manier gebruiken als in hun T1 (Nederlands), bv. SD_F0 en SD_Int. We kunnen daarom de conclusie trekken dat T2-sprekers T1-transfer gebruiken als strategie om vocale emoties te produceren in de T2. Deze strategie zal echter niet voor alle emoties het gewenste resultaat opleveren, bv. niet voor 'verbaasd' en 'sarcastisch'. De resultaten suggereren bovendien dat de T2-sprekers niet automatisch de signalering van emotionele prosodie overnemen van voorbeelden in de doeltaal. Het lijkt er eerder op dat deze gevorderde T2-sprekers van het Chinees een hybride systeem hebben ontwikkeld om emotionele prosodie in de T2 te produceren. Dat hybride systeem benadert de manier waarop moedertaalsprekers van het Chinees vocale emoties uitdrukken, bv. de manier waarop tempo, gemiddelde F0, helling van de F0, compactheid en jitter worden ingezet, maar houdt wat betreft variabiliteit in F0 en intensiteit vast aan wat gebruikelijk is in de moedertaal van de T2-sprekers.

Hoofdstuk 7 onderzoekt de waarneming over en weer van emotionele prosodie die wordt moedertaalluisteraars en door ongeoefende T2-luisteraars (Exp 3). Twintig Chinese en 20 Nederlandse moedertaalluisteraars zonder enige eerdere ervaring met elkaars taal, hadden de opdracht om de zes emotionele prosodische vormen in deze twee talen te identificeren. De resultaten geven aan dat onervaren Nederlandse luisteraars (46% correct) de emotionele prosodie in de onbekende taal (Chinees) even goed konden herkennen als de Chinese moedertaalluisteraars zelf (46% correct); en zij voerden hun taak significant beter als ze dezelfde emotionele prosodieën moesten herkennen in hun moedertaal (Nederlands, 57% correct). De Chinese onervaren luisteraars daarentegen waren alleen in staat om redelijk goed de emotionele prosodie te herkennen als die gesproken was in hun eigen taal (46% correct) maar slaagden er niet in om de vocale emoties boven kans te herkennen als die werd uitgedrukt in de voor hun onbekende taal (Nederlands, 15% correct). Dit resultaat bevestigt het bestaan van het in-groepvoordeel dat ook door andere onderzoekers gerapporteerd is en dat erop neerkomt dat luisteraars emotionele prosodie in het algemeen beter herkennen als die is geproduceerd door sprekers van hun eigen taal dan door sprekers van een onbekende taal. De resultaten suggereren bovendien dat de perceptie van vocale emotie cross-cultureel niet symmetrisch is, zodat de ene groep luisteraars (afhankelijk van taal en/of cultuur) over de hele linie beter is in het identificeren van emotionele prosodie dan de andere groep. Dit ondersteunt op zijn beurt de functionele hypothese die voorspelt dat luisteraars met een toontaalachtergrond in het algemeen minder bedreven zullen zijn in de perceptie van emotionele prosodie dan luisteraars die een niet-toontaal als moedertaal hebben.

Hoofdstuk 8 zet de belangrijkste bevindingen van dit proefschrift nog eens op een rij en probeert dan aan de hand daarvan antwoord te geven op de onderzoeksvragen die in hoofdstuk 1 gesteld zijn. Het hoofdstuk wordt afgesloten met een bespreking van aspecten die in toekomstig onderzoek verbeterd zouden kunnen worden.

