



Universiteit  
Leiden  
The Netherlands

## **When speech becomes emotional: cross-cultural vocal emotion recognition in Dutch and Korean**

Liang, Y.

### **Citation**

Liang, Y. (2025, December 16). *When speech becomes emotional: cross-cultural vocal emotion recognition in Dutch and Korean*. LOT dissertation series. LOT, Amsterdam.  
Retrieved from <https://hdl.handle.net/1887/4285352>

Version: 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/1887/4285352>

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

# Samenvatting

Sinds Darwins "The Expression of the Emotions in Man and Animals" (1872; herdruk in 1998) heeft emotieonderzoek steeds meer aandacht gekregen in verschillende vakgebieden, zoals biologie, psychologie en taalkunde. Darwin stelde dat de productie en perceptie van emoties biologisch bepaald en universeel zijn, terwijl de sociaal-constructivistische theorie stelde dat emoties cultureel en taalkundig geconstrueerd zijn (Harre, 1986). Volgens de dialecttheorie is emotieherkenning fundamenteel universeel, maar kunnen culturele en taalkundige varianten in expressieve stijlen de nauwkeurigheid van de herkenning beïnvloeden (Elfenbein & Ambady, 2002b). Recentelijk groeit de consensus dat cross-culturele emotieherkenning het resultaat is van interacties tussen universele, culturele en taalkundige factoren (Elfenbein, 2013; Elfenbein, Mandal et al., 2002; Mesquita & Frijda, 1992).

Hoewel eerdere studies hebben aangetoond dat vocale emoties over culturen heen herkend kunnen worden, is de vraag nog steeds in hoeverre de productie en perceptie van emoties universeel zijn of beïnvloed worden door cultuur en taal. Dit proefschrift behandelt deze kwestie door te onderzoeken hoe emoties vocaal worden uitgedrukt en waargenomen in twee typologisch en cultureel verschillende talen: Nederlands en Koreaans. Met behulp van het Demo/Koremo-corpus (Broersma et al., 2025) onderzoekt dit proefschrift de herkenning van vocale emoties in verschillende talen, zowel met discrete als dimensionale benaderingen. Er waren drie perceptie-experimenten met luisteraars en één studie gebaseerd op een uitgebreide akoestische analyse van de stimuli, waarbij de volgende aspecten werden onderzocht: 1) de crossculturele en/of crosslinguïstische herkenning van vocale emoties door Nederlandse en Koreaanse luisteraars; 2) de intensiteitsbeoordelingen van vocale emoties door Nederlandse en Koreaanse luisteraars; 3) de relatieve bijdragen van emoties, spreektaal en geslacht aan akoestische parameters; 4) de impact van cultuur en prosodische gelijkenis op de herkenning van vocale emoties door Amerikaans-Engelse en Franse luisteraars.

**Hoofdstuk 1** "Inleiding" presenteert een overzicht van de theoretische en empirische studies naar de productie en perceptie van emoties, waarbij de discrete en dimensionale benaderingen worden besproken. Het introduceert een gebalanceerd "twee-bij-twee"-ontwerp, met sprekers en luisteraars uit twee typologisch verschillende talen: Nederlands en Koreaans. Het omvatte vier basisemoties (angst, boosheid, verdriet, vreugde) en vier niet-basis-emoties (irritatie, opluchting, tederheid, trots), gebalanceerd in opwindend

(hoge versus lage opwindings) en valentie (positief/aangenaam versus negatief/onaangenaam). Om semantische aanwijzingen te vermijden, werden alle emoties geproduceerd op een pseudozin /nuto hɔm sepikɑŋ/, die uitspreekbaar is in het Nederlands en het Koreaans, maar in beide talen geen betekenis heeft. De acht emoties werden uitgedrukt door acht Nederlandse en acht Koreaanse stemacteurs, met vier vrouwen en vier mannen per taalgroep. Dit stelde ons in staat om gendergerelateerde verschillen in prosodische expressie van emoties te bestuderen (Klatt & Klatt, 1990). Elke acteur produceerde twee keer dezelfde emoties, wat resulteerde in een totaal van 256 vertolkingen (8 emoties  $\times$  8 acteurs  $\times$  2 tokens  $\times$  2 talen). Ten slotte schetst dit hoofdstuk de onderzoeksvragen die in elk van de volgende hoofdstukken worden behandeld.

**Hoofdstuk 2** “Onderzoek naar cross-culturele vocale emotieherkenning met een affectief en taalkundig gebalanceerd ontwerp”, onderzoekt de herkenning van vocale Nederlandse en Koreaanse emoties door Nederlandse en Koreaanse luisteraars. Dit hoofdstuk onderzocht 1) of er een in-group voordeel is bij cross-culturele emotieherkenning; 2) of er een verschil is in herkenningsnauwkeurigheid tussen emoties met hoge en lage opwindings; 3) of er een verschil is in herkenningsnauwkeurigheid tussen positieve en negatieve emoties; 4) of er een verschil is in herkenningsnauwkeurigheid tussen basis- en niet-basisemoties.

De resultaten lieten zien dat beide luisteraarsgroepen vocale emoties boven kans herkenden, zelfs in de onbekende taal. Bovendien vertoonden beide luisteraarsgroepen een in-group voordeel bij het herkennen van vocale emoties, zodat luisteraars vocale emoties die in hun moedertaal worden geproduceerd nauwkeuriger herkenden dan die welke in de onbekende taal werden uitgedrukt. Deze bevindingen ondersteunen het idee dat cross-culturele emotieherkenning het resultaat is van een interactie tussen universele en cultuur-/taalspecifieke factoren (Elfenbein, 2013; Elfenbein & Ambady, 2002b). Bovendien onderzocht deze studie de dimensionale effecten van opwindings, valentie en basiciteit op de herkenning van vocale emoties, binnen en tussen culturen. De herkenningsnauwkeurigheid bleek hoger voor lage opwindings, negatieve en basisemoties dan voor hoge opwindings, positieve en niet-basisemoties, zowel binnen als tussen culturen.

**Hoofdstuk 3** “Interpretatie van de intensiteit van vocale emoties tussen culturen” onderzoekt de intensiteitsbeoordelingen van Nederlandse en Koreaanse luisteraars verzameld in Studie 1. Intensiteit is de sterkte van een emotie die door de ontvanger wordt waargenomen (Bänziger & Scherer, 2005; Diener et al., 1985; Larsen & Diener, 1987; Sonnemans & Frijda, 1994). Mensen reageren over het algemeen sterker op emoties met een hogere

intensiteit dan op emoties met een lagere intensiteit. Bovendien geven individuen doorgaans hogere intensiteitsbeoordelingen aan emoties die worden geuit door leden van dezelfde of vergelijkbare cultuur/taalgroep dan door leden van een typologisch andere groep, wat bekend staat als de in-group intensiteitsbias (Kommattam et al., 2019). Dit hoofdstuk onderzocht 1) of accurate trials hogere intensiteitsbeoordelingen krijgen dan onnauwkeurige; 2) of er een in-group bias is voor intensiteitsbeoordelingen tussen culturen; 3) of de drie (dimensionale) binaire splitsingen in opwinding, valentie en basiciteit de intensiteitsbeoordelingen betrouwbaar kunnen voorspellen.

De bevindingen wezen uit dat accurate trials van beide groepen luisteraars hogere intensiteitsbeoordelingen kregen dan inaccurate trials. Boosheid kreeg met name de hoogste intensiteitsbeoordelingen van beide luisteraarsgroepen, zowel in accurate als inaccurate trials. Er werd echter geen in-group bias gevonden in de intensiteitsbeoordelingen. Bovendien waren de intensiteitsbeoordelingen nauw gerelateerd aan opwinding, valentie en basiciteit, zodat de intensiteitsbeoordelingen hoger waren voor emoties met hoge opwinding, negatieve en basisemoties dan voor emoties met lage opwinding, positieve en niet-basisemoties.

**Hoofdstuk 4** “Emoties classificeren op basis van akoestische parameters”, analyseerde 17 akoestische parameters, die werden ingedeeld in vijf categorieën gerelateerd aan: 1) toonhoogte, 2) amplitude, 3) spectrale samenstelling, 4) duur en 5) perturbatie (stemonvastheid). Deze studie onderzocht 1) hoe de akoestische patronen verschillen tussen emoties, sprekerstaal en geslacht; 2) hoe nauwkeurig de herkenning van emoties kan worden voorspeld op basis van akoestische parameters, en in hoeverre de classificatie verbetert wanneer de materialen per taal worden gesplitst; 3) in hoeverre herkenning van emoties d.m.v. machine learning (Support Vector Machine, SVM) overeenkomt met de prestaties van menselijke luisteraars.

De resultaten lieten zien dat elke emotie een uniek akoestisch profiel vertoonde, dat varieerde per taal en geslacht. Vocale emoties kunnen betrouwbaar worden onderscheiden met een SVM-classificator door middel van geoptimaliseerde combinaties van akoestische signalen. De classificatienauwkeurigheid verbeterde wanneer de Nederlandse en Koreaanse gegevens afzonderlijk werden geanalyseerd. De verwarringsmatrices suggereren dat er relevante kwalitatieve verschillen zijn tussen de prestaties van SVM en die van menselijke luisteraars, die nader onderzocht moeten worden.

**Hoofdstuk 5** “Het herkennen van vocale emoties in onbekende talen”, breidt het onderzoek uit naar cross-linguïstische herkenning door Amerikaans-

Engelse en Franse luisteraars die niet bekend waren met het Nederlands en Koreaans. Deze studie onderzoekt de relatieve invloed van de universaliteits-hypothese, culturele nabijheid, linguïstische nabijheid en emotionele dimensies (opwinding, valentie en basiciteit) op emotieherkenning. Bij het herkennen van vocale emoties herkennen luisteraars emoties gemakkelijker in talen die vergelijkbaar zijn met hun eigen taal. Hoewel emotionele dimensies zoals opwinding, valentie en basiciteit de herkenning beïnvloeden, is hun specifieke impact op de nauwkeurigheid niet volledig begrepen. Deze studie onderzoekt hoe universele, culturele, linguïstische en emotionele factoren de perceptie van vocale emoties beïnvloeden. We selecteerden Amerikaans-Engelse en Franse luisteraars omdat Engels, een taal met klemtoontiming, prosodisch vergelijkbaar is met het Nederlands, terwijl Frans, een taal met lettergreep timing, vergelijkbaar is met het Koreaans. Door de herkenning-nauwkeurigheid van deze groepen te vergelijken, willen we de invloed van deze factoren verduidelijken. De studie beoordeelde eerst of beide groepen vocale emoties boven verwachting herkenden. Vervolgens vergeleken we hun nauwkeurigheid met Nederlandse opnames. Daarna evalueerden we of Franse luisteraars beter presteerden dan Amerikaans-Engelse luisteraars met Koreaanse opnames, gezien de prosodische overeenkomsten. Tot slot analyseerden we de effecten van opwinding, valentie en basiciteit op de herkenning van vocale emoties.

De resultaten tonen aan dat zowel Amerikaanse als Franse luisteraarsgroepen vocale emoties boven kans herkenden, wat de universaliteitshypothese steunt (Elfenbein, 2013; Elfenbein & Ambady, 2002a). Bovendien geven de resultaten aan dat mensen emoties nauwkeuriger herkennen wanneer ze hun culturele en taalkundige achtergrond delen met de spreker, een bevinding die in lijn is met Elfenbein en Ambady (2003a).

**Hoofdstuk 6** “Conclusies en discussie” herhaalt de onderzoeksvragen van elk hoofdstuk en vat de belangrijkste bevindingen van elke studie samen met een discussie en een integratie van de resultaten. Dit hoofdstuk behandelt ook de beperkingen van de empirische studies. Tot slot worden de volgende vijf nieuwe bevindingen en bijdragen aan de bestaande literatuur over emotieonderzoek genoemd.

Ten eerste presenteert onze studie een evenwichtig “twee-bij-twee” gekruist experimenteel design, met sprekers en luisteraars uit twee typologisch verschillende talen en culturen. Dit design maakt een systematische vergelijking mogelijk van herkenningsspatronen binnen en tussen culturen, met name om het in-group voordeel in emotieherkenning te onderzoeken. We repliceerden de eerdere bevinding dat luisteraars vocale emoties boven kans

herkenden, zelfs in een onbekende taal, wat consistent is met de universaliteitshypothese. Bovendien ontdekten we dat beide luisteraarsgroepen een in-group voordeel vertoonden, waardoor ze vocale emoties nauwkeuriger herkenden wanneer ze in hun moedertaal werden uitgesproken dan wanneer ze in de onbekende taal werden uitgesproken.

Ten tweede hebben we zowel discrete als dimensionale benaderingen gebruikt, met vier basis- en vier niet-basisemoties, gebalanceerd naar opwinding en valentie, wat een uitgebreider begrip opleverde van de mechanismen die ten grondslag liggen aan emotieherkenning. Deze studie onderzocht niet alleen de herkenningsnauwkeurigheid van elk van de acht emoties categorisch, maar ook de effecten van drie emotionele dimensies (opwinding, valentie en basisgevoel) op emotieherkenning, binnen en tussen culturen. De herkenningsnauwkeurigheid bleek hoger voor lage-opwinding, negatieve en basisemoties dan voor hoge-opwinding, positieve en niet-basisemoties, zowel binnen als tussen culturen. Voor zover wij weten is dit de eerste studie die de herkenningsnauwkeurigheid van lage-opwinding en hoge-opwinding emoties binnen en tussen culturen direct heeft vergeleken. Deze bevinding verklaart de rol van opwinding bij emotieherkenning.

Ten derde onderzochten we intensiteitsbeoordelingen in verschillende culturen. Intensiteit, als belangrijke dimensie van emoties, heeft relatief weinig aandacht gekregen in eerder onderzoek. We vonden geen in-group bias in intensiteitsbeoordelingen. De resultaten lieten wél zien dat intensiteitsbeoordelingen voornamelijk werden bepaald door het type emotie in plaats van door culturele of taalkundige factoren. Met behulp van de dimensionale benadering onderzochten we de rol van opwinding, valentie en basiciteit in intensiteitsbeoordelingen. We ontdekten dat intensiteitsbeoordelingen over het algemeen hoger waren voor emoties met hoge opwinding, negatieve emoties en basisemoties dan voor emoties met lage opwinding, positieve emoties en niet-basisemoties.

Ten vierde biedt onze studie een uitgebreide analyse van 17 akoestische parameters in twee typologisch verschillende talen, waaronder een ondervertegenwoordigde niet-Indo-Europese taal: Koreaans. We ontdekten dat verschillende vocale emoties universele en emotiespecifieke akoestische patronen vertonen. Met name toonhoogte, intensiteit en spreek snelheid kunnen vocale emoties succesvol differentiëren. We ontdekten ook dat laryngale parameters genuanceerde onderscheidingen tussen vocale emoties mogelijk maken. Bovendien vertonen vocale emoties taalafhankelijke akoestische patronen. De articulatiesnelheid was opmerkelijk hoger in het Koreaans dan in het Nederlands, en er waren geen spraakpauzes in het

Koreaans. Bovendien gebruikten vrouwen en mannen verschillende akoestische eigenschappen, onder andere doordat (relatieve) toonhoogte- en amplitudegerelateerde akoestische parameters over het algemeen hoger zijn voor vrouwen dan voor mannelijke acteurs. Op basis van de akoestische parameters onderzochten we of vocale emoties nauwkeurig kunnen worden geclassificeerd door middel van machine learning (SVM-modellen) en vergeleken we de prestaties van machinale classificatoren voor Nederlands en Koreaans in zowel in-group- als out-group-omstandigheden. De resultaten wezen uit dat classificatienauwkeurigheid betrouwbaar kan worden voorspeld uit combinaties van akoestische parameters. Opvallend is dat machinale classificatie, net als menselijke luisteraars, beter presteerde in de in-group-conditie dan in de out-group-conditie voor zowel Nederlanders als Koreanen, hoewel de classificatiepercentages in beide condities boven kans lagen.

Ten slotte onderzochten we cross-culturele/linguïstische herkenning van vocale emoties door Amerikaans-Engelse en Franse luisteraars, die geen kennis hadden van Nederlands of Koreaans. Deze studie breidt voor de bestaande literatuur uit over cross-culturele/linguïstische herkenning van emoties door de bijdragen van universaliteit, culturele nabijheid, prosodische nabijheid en emotionele dimensies (opwinding, valentie en basiciteit) aan cross-culturele herkenning van vocale emoties te testen. In overeenstemming met de Universaliteitshypothese identificeerden beide luisteraarsgroepen vocale emoties over het algemeen boven kans, zelfs in een onbekende taal. We vonden echter geen algemeen effect van culturele afstand op emotieherkenning, aangezien de herkenningsnauwkeurigheid varieerde per emotie, noch vonden we ondersteuning van het effect van prosodische afstand op vocale emotieherkenning over culturen/talen heen, aangezien de Franse luisteraars het met de Koreaanse opnames niet beter deden dan de Amerikaans-Engelse luisteraars. Belangrijker nog, we ontdekten dat hoewel opwinding, valentie en basiciteit de herkenning van vocale emoties beïnvloeden, sommige emoties deze algemene patronen schonden. Hoewel emotionele dimensies een bruikbaar raamwerk bieden om de nauwkeurigheid van herkenning te begrijpen, verklaren zij de variatie in emotieherkenning tussen culturen en talen vooralsnog onvoldoende.