

**Multimodale Coördinatie in L1 en L2: De Relatie Tussen Representerende
Gebaren en Spraakverstoringen**

Nina Jetten

Studentnummer: s4784448

Afdeling Psychologie, Rijksuniversiteit Groningen

PSB3A-BT15: Bachelor These

Begeleider: dr. Lisette de Jonge-Hoekstra

Tweede beoordelaar: prof. dr. Henk A. L. Kiers

In samenwerking met: Melina Bardi, Natalia Kijak, Samuel Martire, Lynet Reinds en

Vladimír Vursta

21 april 2025

Een scriptie is een proeve van bekwaamheid voor studenten. De goedkeuring van de scriptie is het bewijs dat de student over voldoende onderzoeks- en rapportagevaardigheden beschikt om af te studeren, maar biedt geen garantie voor de kwaliteit van het onderzoek en de resultaten van het onderzoek als zodanig, en de scriptie is daarom niet per se geschikt als academische bron om naar te verwijzen. Als u meer wilt weten over het in deze scriptie besproken onderzoek en de daarop gebaseerde publicaties waarnaar u zou kunnen verwijzen, neem dan contact op met de genoemde begeleider.

Abstract

Communication between people is more than just words and occurs both verbally and non-verbally. There is substantial evidence that gestures are closely linked to speech, indicating a strong interrelation. This study investigates the extent to which representational gestures are associated with speech disfluency in individuals speaking their first (L1) or second (L2) language. Participants were paired into dyads across three different conditions: L1-L1, L1-L2, and L2-L2. The dyads were filmed while discussing various statements. Speech and gestures were coded afterwards following the M3D guidelines. We found no significant differences between the three conditions in the frequency of representational gestures, which is not in line with previous research. Nor did we find an effect of condition on representational gestures during speech disfluency, which contradicts the *lexical retrieval hypothesis*. An additional paired-samples t-test showed that significantly more gestures were produced during fluent speech than during disfluent speech. This aligns with earlier research indicating that gestures tend to cease when speech ceases. The current study represents a first step in demonstrating this pattern, but further research is needed to clarify these findings, with a focus on differences between L1 and L2 speakers.

Keywords: representational gestures, L1, L2, speech disfluency

Samenvatting

Communicatie tussen mensen is meer dan alleen woorden en vindt zowel verbaal als non-verbaal plaats. Er is aanzienlijk bewijs dat gebaren nauw verbonden zijn met spraak, wat wijst op een sterke onderlinge samenhang. Deze studie onderzoekt in welke mate de representerende gebaren samenhangen met spraakverstoringen bij mensen die in hun eerste (L1) of tweede (L2) taal spreken. Participanten werden ingedeeld in tweetallen in drie verschillende condities: L1-L1, L1-L2 en L2-L2. De tweetallen werden gefilmd, terwijl zij discussieerden over stellingen. Spraak en gebaren werden na afloop gecodeerd volgens de richtlijnen van M3D. We vonden geen significante verschillen tussen de drie condities in het aantal representerende gebaren, wat niet in lijn is met eerder onderzoek. Ook vonden we geen effect van condities op representerende gebaren tijdens spraakverstoringen. Dit is in strijd met *lexical retrieval hypothesis*. Een aanvullende analyse toonde dat er significant meer gebaren werden gemaakt tijdens vloeiende spraak dan tijdens spraakverstoringen. Dit is in lijn met eerder onderzoek, waaruit bleek dat gebaren stopten wanneer spraak stopte. Het huidige onderzoek is een eerste stap om dit aan te tonen, maar vervolgonderzoek is noodzakelijk om dit beter te verklaren, met een focus op de verschillen tussen L1- en L2-sprekers.

Trefwoorden: representerende gebaren, L1, L2, spraakverstoringen

Multimodale Coördinatie in L1 en L2: De Relatie Tussen Representerende Gebaren en Spraakverstoringen

Communicatie tussen mensen is veel meer dan alleen woorden. Het is een multimodaal proces, waarbij er naast spraak gebruik wordt gemaakt van taal-gerelateerd fysiek gedrag, zoals handgebaren of hoofdbewegingen. Deze vormen van non-verbale communicatie – zoals gebaren, gezichtsuitdrukkingen en lichaamshouding – versterken en vergemakkelijken de communicatie tussen gesprekspartners (Gullberg, 2022). Gebaren worden gedefinieerd als zichtbare acties van een lichaamsdeel die gebruikt worden als een uiting of als onderdeel van een uiting (Kendon, 2004). Er is aanzienlijk bewijs dat gebaren systematisch en nauw verbonden zijn met taal bij zowel spraakproductie als spraakbegrip (Gullberg, 2022). Dit wijst op een sterke onderlinge samenhang tussen spraak en gebaren. Zo ontwikkelen ze gelijktijdig bij kinderen (Gullberg et al., 2008). Dit is ook te zien bij communicatiestoornissen – zoals stotteren (Mayberry & Jaques, 2000) – waarbij kinderen moeite hebben met zowel spraak als met gebaren. Gebaren zijn dus een essentieel onderdeel van de communicatie tussen mensen, wat het relevant maakt om het verband tussen spraak en gebaren te onderzoeken. Daarnaast is het door globalisering steeds belangrijker geworden om meerdere talen te beheersen. Zo leren kinderen van jongs af naast hun moedertaal ook andere talen. Dit maakt het interessant om de verschillen te onderzoeken tussen mensen die hun eerste (L1) of tweede (L2) taal spreken. In deze studie zal verder worden ingegaan op de mate waarin de frequentie van gebaren samenhangt met spraakverstoringen bij het spreken van een L1 of L2.

1.1 Spreken in L2

Spreken in een L2 verschilt op meerdere manieren van spreken in een L1. Zo spreken de meeste mensen minder vloeiend in hun L2, alhoewel de vloeiendheid van de L2 over tijd toeneemt (De Jong, 2016). De meest voor de hand liggende reden voor de minder vloeiende spraak in L2 is een gebrek aan taalvaardigheid, wat inhoudt dat een persoon onvoldoende kennis

over de L2 heeft en te weinig oefening heeft gehad met het spreken van de taal (Bergmann et al., 2015). Een tweede reden voor minder vloeiende spraak in L2 is de invloed van de L1. Bij het spreken van een L2 kan er sprake zijn van taal co-activatie (Bergmann et al., 2015). Taal co-activatie is de activatie van representaties die gebruikt worden in de L1 tijdens het spreken van de L2. Deze automatische activatie is moeilijk te onderdrukken, wat kan zorgen voor interferentie bij het spreken van de L2 (Bergmann et al., 2015). Omdat mensen vaak minder vloeiend spreken in hun L2, kunnen gebaren een belangrijke rol spelen in het ondersteunen van de communicatie (Bergmann et al., 2015), waar later dieper op in wordt gegaan.

1.2 Soorten gebaren

Gebaren kunnen worden geclassificeerd aan de hand van hun semantische functie. Hierbij kan het classificatiesysteem van McNeill (1992) worden gebruikt om de verschillende typen gebaren te onderscheiden. McNeill (1992) maakte onderscheid tussen emblematische, deictische, ritmische, iconische en metaforische gebaren. Emblematische gebaren hebben een vaste betekenis en functioneren onafhankelijk van spraak, zoals het opsteken van een duim om aan te geven dat er ergens mee wordt ingestemd (Wagner et al., 2013). Deictische gebaren wijzen naar objecten of locaties in een ruimte. Deze gebaren kunnen zowel concreet – bijvoorbeeld het wijzen naar een persoon in de ruimte – als abstract – bijvoorbeeld naar een plek achter het lichaam wijzen om naar het verleden te verwijzen – zijn (Rohrer et al., 2023). Ritmische (“beat”) gebaren zijn simpele en snelle handbewegingen zonder inhoudelijke betekenis, die enkel ondersteunend zijn voor de structuur van spraak. Iconische gebaren visualiseren concrete objecten of acties, zoals de vorm of beweging van een object (Wagner et al., 2013). Een voorbeeld van een iconisch gebaar is wanneer iemand de lengte van een voorwerp nabootst met zijn/haar handen om aan te geven hoe lang het voorwerp in werkelijke grootte is. Metaforische gebaren zijn iconische gebaren die abstracte concepten weerspiegelen (Wagner et al., 2013). Een voorbeeld van een metaforisch gebaar is wanneer iemand met

zijn/haar handen op een neer beweegt alsof het een weegschaal is, om zo een afweging of innerlijk dilemma uit te drukken. Iconische en metaforische gebaren kunnen gezamenlijk worden geclassificeerd als representerende gebaren (McNeill, 1992). In deze studie wordt er gefocust op representerende gebaren.

1.3 Representerende gebaren en spraakverstoringen in L1 en L2

Representerende gebaren zijn hand- of armgebaren die een concept visueel weergeven (Nicoladis et al., 2018) en worden gebruikt om de betekenis van spraak te ondersteunen of te verduidelijken. Een voorbeeld van een representerend gebaar is wanneer iemand een cirkel vormt met zijn/haar handen om een voetbal na te bootsen. Dit kan spraak verduidelijken wanneer de spreker het woord ‘voetbal’ niet kan ophalen of als de gesprekspartner het woord ‘voetbal’ niet kent. Een belangrijke functie van representerende gebaren is dat ze helpen bij het ophalen van woorden uit het geheugen (*lexical retrieval*). Volgens de *lexical retrieval hypothesis* activeren gebaren de sensomotorische kenmerken van op te halen woorden, waardoor ze gemakkelijker worden opgehaald en geproduceerd (Krauss, 1998; Arslan et al., 2024). Deze hypothese ondersteunt het idee dat er bij spraakverstoringen meer gebruik wordt gemaakt van gebaren dan bij vloeiende spraak.

Vloeiende spraak is gedefinieerd als spraak zonder onnatuurlijke aarzelingen (De Jong, 2016). De vloeiendheid van spraak hangt af van de kennis over de taal, zoals een grote woordenschat en het snel kunnen ophalen van woorden. Bij spraakverstoringen zijn deze factoren afwezig (Arslan et al., 2024). Spraakverstoringen zijn tijdelijke onderbrekingen in spraak zonder dat er een inhoudelijke bijdrage wordt geleverd aan de uiting (MacLay & Osgood, 1959; Tree, 1995). Deze onderbrekingen kunnen in verschillende vormen voorkomen, zoals een ongevolde pauze (stilte), een gevulde pauze (bijvoorbeeld ‘uh’), herhalingen of zelfverbeteringen, waarbij een eerdere uiting wordt gecorrigeerd (MacLay & Osgood, 1959). In L2 blijken spraakverstoringen vaker voor te komen dan in L1. Dit wordt bevestigd door het

onderzoek van Bergmann et al. (2015), waarin dit wordt verklaard doordat mensen een lagere taalvaardigheid hebben in L2 in vergelijking met L1. Een manier om spraakverstoringen te compenseren, zou mogelijk het gebruiken van gebaren kunnen zijn.

Mensen gebruiken meer gebaren als ze spreken in hun L2 dan in L1 (Arslan et al., 2024). Dit bleek ook uit onderzoek van Hadar et al. (2001), waar het gebruik van ‘ideational gestures’ – een synoniem voor representerende gebaren – werd onderzocht bij mensen met Hebreeuws als L1 en Engels als L2. In het Engels (L2) werd er meer gebruik gemaakt van ‘ideational gestures’ dan in het Hebreeuws (L1), wat suggereert dat gebaren ondersteunend zijn in het proces van woordophaling bij het spreken van een L2.

1.4 Frequentie van gebaren en spraakverstoringen

Als mensen daadwerkelijk gebaren gebruiken om te compenseren voor moeilijkheden in hun spraak, dan zou dit betekenen dat gebaren voornamelijk voorkomen tijdens spraakverstoringen (Graziano & Gullberg, 2018). Graziano en Gullberg (2018) onderzochten deze aanname door tweetaligen te vergelijken met mensen die een taal aan het leren waren (kinderen en volwassen L2-leerlingen). Uit hun onderzoek bleek dat sprekers in alle groepen vooral gebaren gebruikten tijdens vloeiende spraak en zelden tijdens spraakverstoringen. Deze bevindingen suggereren dat gebaren stoppen wanneer spraak stopt, wat niet in lijn is met het onderzoek van Arslan et al. (2024), Hadar et al. (2001) en de *lexical retrieval hypothesis*. Arslan et al. (2024) stelden dat gebaren tijdens spraakverstoringen worden gebruikt als mensen in staat zijn hun eigen fouten te herkennen. Dit wijst mogelijk op een verband tussen representerende gebaren en taalvaardigheid bij het behalen van voordelen uit gebaren bij spraakverstoringen (Arslan et al., 2024). Er lijkt tot op heden geen consensus te zijn over de relatie tussen spraakverstoringen en de frequentie van gebaren. Daarom is het belangrijk hier onderzoek naar te doen.

1.5 Huidige studie

Deze studie onderzoekt in welke mate de frequentie van representerende gebaren samenhangt met spraakverstoringen bij mensen die hun eerste (L1) of tweede (L2) taal spreken. Om dit te onderzoeken wordt er een experiment uitgevoerd, waarbij participanten in tweetallen discussiëren over stellingen. Participanten worden ingedeeld in drie groepen, waarbij in de eerste groep beide participanten in hun L1 spreken, in de tweede groep beide participanten in hun L2 spreken en in de derde groep de ene participant in zijn/haar L1 spreekt en de ander in zijn/haar L2. Op basis van eerder onderzoek zijn er drie hypothesen. In lijn met eerder onderzoek van Hadar et al. (2001) en Arslan et al. (2024), waaruit bleek dat mensen meer gebaren gebruik in hun L2 dan in hun L1, is de eerste hypothese dat participanten in de L1-L1 conditie minder representerende gebaren gebruiken dan participanten in de L2-L2 en L1-L2 conditie. De tweede hypothese is dat er in de L1-L2 conditie minder representerende gebaren voorkomen dan in de L2-L2 conditie. Mogelijk komt dit doordat het lagere aantal representerende gebaren van de L1-spreker compenseert voor het hogere aantal van de L2-spreker. Tot slot is de derde hypothese dat tijdens spraakverstoringen in de L2-L2 conditie meer representerende gebaren worden gemaakt dan in de L1-L1 en L1-L2 conditie. Dit is in lijn de *lexical retrieval hypothesis* en de bevinding van Arslan et al. (2024) dat gebaren tijdens spraakverstoringen worden gebruikt als mensen in staat zijn hun eigen fouten te herkennen. Tegelijkertijd staat deze hypothese haaks op het onderzoek van Graziano & Gullberg (2018), die juist vonden dat gebaren stoppen wanneer spraak stopt. Dit verschil in verwachting komt door de aanname dat L2-sprekers gebaren gebruiken om de communicatie te ondersteunen en te compenseren voor taalproblemen, waar L1-sprekers deze extra ondersteuning niet nodig hebben en dan mogelijk pauzeren in zowel spraak als gebaren. Door deze hypothesen te testen wordt er dieper inzicht verkregen in de rol van representerende gebaren bij spraakverstoringen door de condities L1-L1, L1-L2 en L2-L2 te vergelijken.

2.0 Methode

2.1 Participanten

In totaal namen 78 participanten deel aan het onderzoek, bestaande uit 39 tweetallen. Van 66 participanten werd de data gecodeerd ($M_{age} = 20.14$, $SD_{age} = 1.84$). Alle participanten waren bachelor studenten die via een gemakssteekproef werden geworven. Ter compensatie voor de deelname aan het onderzoek, ontvingen de participanten studiepunten die vereist waren voor het behalen van een eerstejaars vak Psychologie. Hoewel de studenten verplicht waren een bepaald aantal studiepunten te behalen, hadden ze de vrijheid om zelf te kiezen aan welk experiment zij wilden deelnemen.

Participanten voerden het gesprek in het Nederlands of in het Engels. Andere talen die deelnemers aangaven te spreken, waren Russisch, Slowaaks, Fins, Roemeens, Chinees, Japans, Spaans, Pools, Hindi, Bengaals, Arabisch, Frans, Italiaans, Zweeds, Turks, Tsjechisch, Amerikaanse Gebarentaal, Fries, Kroatisch, Thais en Grieks. Het onderzoek werd goedgekeurd door de Ethische Commissie van de Rijksuniversiteit Groningen (code: PSY-2425-S-0110) en alle participanten gaven schriftelijk geïnformeerde toestemming.

2.2 Materialen

Om de dataverzameling in een gecontroleerde omgeving uit te voeren, vond het experiment plaats in een speciaal voor gedragswetenschappelijk onderzoek ingerichte ruimte van de faculteit Gedrags- en Maatschappijwetenschappen. Twee stoelen stonden in een hoek van 90 graden dicht bij elkaar opgesteld om communicatie tussen deelnemers te vergemakkelijken. De deelnemers werden gefilmd met een JVC-camera en een Sony-camera, die beiden bevestigd waren op verstelbare statieven. Deze statieven waren zorgvuldig geplaatst, zodat alle gebaren volledig zichtbaar waren tijdens het gesprek. Een kamerscherm zorgde ervoor dat de deelnemers de onderzoekers niet konden zien, zowel tijdens de testafname als tijdens de experimentele taak. Dit werd gedaan om afleiding te minimaliseren en een zo

natuurlijk mogelijke interactie tussen deelnemers te waarborgen. Eventuele omgevingsgeluiden werden later verwijderd uit de opnames met behulp van de audio-editor Audacity (Audacity Team, 2014). De geluidssterkte van de spraak werd in een aangepast R-script – ontwikkeld door de begeleidster – gebruikt om het begin en het einde van spraaksegmenten te bepalen.

Er waren twee computers beschikbaar in de ruimte om het geïnformeerde toestemmingsformulier, de demografische vragenlijst en de C-test in te vullen. Om de taalvaardigheid van de participanten te meten, werden er vier verschillende versies van de C-test gebruikt: een Engelse (Gilmore, 2013), Nederlandse (Freeborn & Andringa, 2023), Franse (Coleman et al., 2017a) en Duitse (Coleman et al., 2017b) versie¹. Elke paragraaf bevatte meerdere open plekken waar letters moesten worden ingevuld, zoals: “You wa ____ in t ____ morning, ru ____ to t ____ window a ____ take a de ____ breath. Do ____!” (Gilmore, 2013). De taalvaardigheidsscore van de deelnemers werd bepaald aan de hand van het percentage correct ingevulde open plekken.

Daarnaast werden enkele vooraf opgestelde stellingen gebruikt om het debat op gang te brengen, die ongeacht de voertaal van het gesprek in het Engels werden gesteld, omdat niet elke onderzoeker Nederlands sprak. Deze stellingen waren bijvoorbeeld: ‘Which came first? The chicken or the egg?’ en ‘Would you rather be very intelligent or very funny?’. De stellingen waren eerder getest in een proefonderzoek van een masterstudent (zie Bijlage A).

Tot slot werden de ondertekende toestemmingsformulieren, de ingevulde demografische vragenlijsten en de C-testen geüpload in Qualtrics (Qualtrics XM – Experience Management Software, 2025). De onderzoekers volgden een trainingsprogramma voor gebarenannotatie (Tütüncübasi et al., 2023) voordat zij de gegevens codeerden in ELAN (ELAN, 2024). De gegevens werden later samengevoegd in één databestand met behulp van een R-script en geanalyseerd in JASP (JASP, 2025). Een meer gedetailleerde beschrijving van

¹ Uiteindelijk zijn alleen de Nederlandse en Engelse versie van de C-test ingevuld door participanten.

het gebruik van deze programma's volgt in de secties dataverwerking en codering en statistische analyse.

2.3 Procedure

Het experiment werd uitgevoerd in tweetallen. Bij aankomst werden de participanten geïnformeerd over het verloop van het experiment en het algemene doel van de studie. Participanten gaven via een toestemmingsformulier in Qualtrics toestemming voor het opnemen van het experiment. Door het formulier te ondertekenen, stemden ze in met het opnemen en verwerken van hun geschreven, video- en audiomateriaal. Om een natuurlijke productie van gebaren tijdens het gesprek te waarborgen, werd niets verteld over gebaren of spraak. Daarna vulden de deelnemers een online demografische vragenlijst in. Een onderzoeksleider bepaalde aan de hand van de eerste en tweede taal van elke deelnemer in welke taal het tweetal zou communiceren, en vulde dit in de vragenlijst in. Ook werd de zitpositie van elke deelnemer (links of rechts in het beeld) door de onderzoeker genoteerd. Vervolgens maakten de deelnemers een C-test in de taal waarin het experiment werd uitgevoerd om hun taalvaardigheid te meten. Hiervoor kregen ze maximaal acht minuten. Als de test niet binnen deze tijd werd voltooid, werd er gevraagd om een laatste woord op te schrijven en de onvoltooide test in te leveren.

Na de C-test namen beide participanten de aangewezen plaatsen in en begon het daadwerkelijke experiment. De onderzoeker las de stellingen voor, waarover de participanten dienden te discussiëren. Ze werden geïnstrueerd om de voor- en nadelen van elke stelling te bespreken totdat ze tot een consensus kwamen. Zodra ze aangaven een consensus te hebben bereikt, werd de volgende stelling voorgelezen.

De discussie duurde maximaal tien minuten, ongeacht hoeveel stellingen besproken werden. Als de tien minuten verstreken waren terwijl er nog werd gesproken, mochten ze hun laatste discussie afronden. Tot slot werden de participanten bedankt voor hun deelname, kregen

ze een korte uitleg over het onderzoek, mochten ze vragen stellen en vond er een zogenoemde ‘debriefing’ plaats. Het gehele experiment duurde ongeveer 30 minuten per tweetal.

2.4 Dataverwerking en codering

De onderzoekers codeerden de gebaren in de video’s om de gegevens te analyseren met behulp van ELAN (ELAN, 2024). Elke onderzoeker codeerde onafhankelijk maximaal zes video’s. Nederlandse onderzoekers codeerden de Nederlandstalige video’s, terwijl internationale onderzoekers de Engelstalige video’s codeerden. Voorafgaand aan het coderen doorliepen de onderzoekers een uitgebreide M3D-trainingsmodule (Tütüncübasi et al., 2023), die richtlijnen bood voor het gebruik van ELAN en het onderscheiden van verschillende soorten gebaren. Naast het bekijken van schriftelijke instructies en trainingsvideo’s, maakten de onderzoekers oefeningen om vertrouwd te raken met het programma.

Er werd een bijeenkomst gehouden om de interbetrouwbaarheid te bespreken. Een aangepaste sjabloon van het M3D-trainingsmodule werd gebruikt om consistente codering te garanderen. Het sjabloon omvatte: G-units, G-phases (alleen de ‘stroke’), ‘self-adaptors’, niet-referentiële gebaren, metaforische gebaren, iconische gebaren, abstracte deictische gebaren, concrete deictische gebaren en emblemen (Rohrer et al., 2023). Het coderen van G-units hield in dat segmenten waarin deelnemers gebaren maakten werden geïdentificeerd; het begin van een G-unit werd genoteerd wanneer de handen een rustpositie verlieten en het einde wanneer ze terugkeerden naar rust. Voor het identificeren van de ‘stroke’ – de fase van het gebaar die de belangrijkste informatie overbrengt (Rohrer et al., 2023) – moesten de onderzoekers het meest prominente en betekenisvolle deel van de G-unit bepalen. Vervolgens werd voor elke ‘stroke’ het type gebaar gecodeerd, zoals een niet-referentiële gebaren of een ‘self-adaptor’. Niet-referentiële gebaren zijn gebaren die de inhoud van spraak niet visueel weergeven en daardoor geen betekenis dragen (Rohrer et al., 2023). ‘Self-adaptors’ zijn gebaren waarbij het eigen

lichaam wordt aangeraakt, bijvoorbeeld krabben of de nek masseren (Wagner et al., 2013). Elke categorie werd afzonderlijk gecodeerd voor de linker- en rechterdeelnemer.

Naast de gebaren werd ook de spraak gecodeerd. Eerst werd er eventueel achtergrondgeluid automatisch uit de audio verwijderd met behulp van Audacity (Audacity Team, 2014). Vervolgens werden met een aangepast R-script – ontwikkeld door de begeleidster – de start- en stoptijden van de spraak automatisch bepaald op basis van de geluidssterkte in het opgenomen audiobestand. Wanneer een pauze tussen spraaksegmenten korter was dan 250 ms, werden deze segmenten als één segment beschouwd. De hieruit voortgekomen csv-bestanden werden geïmporteerd in ELAN. Om aan te geven wie er sprak in elk segment, pasten de onderzoekers handmatig de annotatie aan naar part.R (rechter participant), part.L (linker participant) of exp.leader (onderzoeker). Wanneer het R-script er niet in slaagde om spraaksegmenten van beide deelnemers te onderscheiden, werden deze handmatig opgesplitst door de onderzoekers. Wanneer twee deelnemers onverstaanbaar door elkaar praatten, werd de annotatie “part.R.part.L” aan dat segment toegekend om aan te geven dat beiden participanten aan het woord waren.

Tot slot werd met een ander aangepast R-script de data uit de demografische vragenlijst, de C-test en de gecodeerde ELAN-bestanden samengevoegd tot één eindbestand dat verdere statistische analyse van de onderzoeksvragen mogelijk maakte. De analyse van de variabelen die relevant zijn voor de onderzoeksvraag van deze bachelor scriptie wordt hierna beschreven.

2.5 Statistische analyse

Om de verschillende hypothesen te toetsen, zijn er statistische analyses uitgevoerd met behulp van JASP. De hypothese dat proefpersonen in de L1-L1 conditie minder representerende gebaren gebruiken dan proefpersonen in de L1-L2 en L2-L2 conditie en de hypothese dat proefpersonen in de L1-L2 conditie minder representerende gebaren gebruiken dan proefpersonen in de L2-L2 conditie, zijn getoetst met een one-way ANOVA. Aanvullend hierop

is er getoetst of er verschillen zijn tussen de condities L1-L1/L1-L2, L1-L1/L2-L2 en L1-L2/L2-L2 door deze condities te vergelijken met een Games-Howell post-hoc multiple comparisons analyse. Er is voor deze post-hoc test gekozen, vanwege de ongelijke groepsgroottes. Ook is er voor de variabele representerende gebaren gecontroleerd op outliers. Waarden die meer dan drie standaarddeviaties zouden afwijken van het gemiddelde, zouden niet worden meegenomen in de analyse. Dit was echter niet het geval. De assumptie van normaliteit is getoetst met de Shapiro-Wilk test. Als deze assumptie werd geschonden, werd er een Kruskal-Wallis test uitgevoerd – een non-parametrisch alternatief voor een ANOVA om drie onafhankelijke groepen te kunnen vergelijken.

De derde hypothese dat er tijdens spraakverstoringen in de L2-L2 conditie meer representerende gebaren worden gemaakt dan in L1-L1 en L1-L2, werd wederom getoetst met een one-way ANOVA, waarbij de conditie de onafhankelijke variabele was en het aantal representerende gebaren tijdens spraakverstoringen de afhankelijke variabele. Ook voor deze hypothese is er gecontroleerd of de assumptie van normaliteit werd geschonden en werd er – indien die werd geschonden – overgestapt op de Kruskal-Wallis test. Er werd opnieuw gecontroleerd op outliers van meer dan drie standaarddeviaties boven het gemiddelde, maar deze werden niet gevonden.

3.0 Resultaten

De mate waarin de frequentie van gebaren samenhangt met spraakverstoringen bij het spreken van een L1 of L2 is getoetst aan de hand van verschillende analyses. De analyse is uitgevoerd met de data van 66 proefpersonen. In totaal hebben 78 studenten in tweetallen deelgenomen aan het experiment, maar om pragmatische redenen (tijd) zijn van 66 proefpersonen de gebaren gecodeerd. Het aantal proefpersonen per conditie is ongelijk: de L1-L1 conditie bevat 18 proefpersonen, de L1-L2 conditie 10 en de L2-L2 conditie 38.

3.1 Verschillen in het aantal representerende gebaren tussen de condities

In Tabel 1 is te zien dat de assumptie van normaliteit werd geschonden voor het aantal representerende gebaren. Een overzicht van de verdeling per conditie is weergegeven in Figuur 1 (zie ook Figuur 1 in Bijlage B). De Shapiro-Wilk test voor zowel de L1-L1 ($p = .002$) als de L2-L2 conditie ($p < .001$) was significant, wat duidt op een afwijking van normaliteit. Om deze reden is er gekozen voor een Kruskal-Wallis test om te toetsen of er significante verschillen zijn tussen de condities.

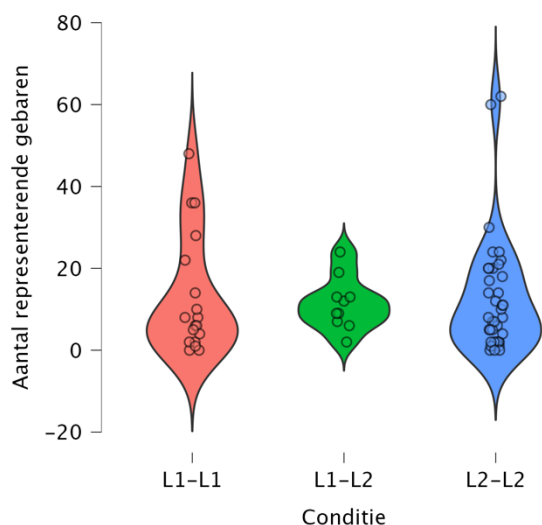
Tabel 1

Descriptives representerende gebaren per conditie

	Aantal representerende gebaren		
	L1-L1	L1-L2	L2-L2
N	18	10	38
Mediaan	7.000	10.500	8.000
IQR	17.500	5.500	17.500
Shapiro-Wilk	0.815	0.954	0.758
P-waarde van Shapiro-Wilk	0.002	0.713	< .001
Minimum	0	2	0
Maximum	48	24	62

Figuur 1

Vioolplot representerende gebaren per conditie



Bij het toetsen van een verschil in het aantal representerende gebaren tussen de condities, toonde een Kruskal-Wallis test geen significant effect van condities op het aantal representerende gebaren van $H(2) = .533, p = .766$. De Games-Howell post-hoc analyse (Tabel 2) toonde geen significante verschillen tussen de condities L1-L1 vs. L1-L2 ($M = 1.711, SE = 3.991, p = .904$), L1-L1 vs. L2-L2 ($M = .532, SE = 4.139, p = .991$) en L1-L2 vs. L2-L2 ($M = -1.179, SE = 3.179, p = .922$). Gezamenlijk tonen deze analyses dat er geen verschil lijkt te zijn in het aantal representerende gebaren tussen de verschillende condities. Dit is niet in lijn met de hypothesen dat proefpersonen in de L1-L1 conditie minder representerende gebaren gebruiken dan proefpersonen in de L1-L2 en L2-L2 conditie en dat proefpersonen in de L1-L2 conditie minder representerende gebaren gebruiken dan proefpersonen in de L2-L2 conditie.

Tabel 2

Games-Howell multiple comparisons L1-L1, L1-L2, L2-L2

Vergelijking	Verskil in gemiddelde	SE	t	df	P _{Tukey}
L1-L1 - L1-L2	1.711	3.991	0.429	25.143	0.904
L1-L1 - L2-L2	0.532	4.139	0.129	32.698	0.991
L1-L2 - L2-L2	-1.179	3.071	-0.384	33.624	0.922

Note. Resultaten zijn gebaseerd op niet-gecorrigeerde gemiddelden.

3.2 Representerende gebaren tijdens spraakverstoringen

In Tabel 3 is te zien dat de assumptie van normaliteit werd geschonden voor het aantal representerende gebaren tijdens spraakverstoringen. Een overzicht van de verdeling is te zien in Figuur 2. De Shapiro-Wilk test toonde een significant resultaat van $p < .001$ voor elke conditie, wat duidt op een afwijking van normaliteit. Om deze reden is er gekozen voor een Kruskal-Wallis test om te toetsen of er significante verschillen zijn tussen de condities.

Bij het toetsen van een verschil in het aantal representerende gebaren tijdens spraakverstoringen tussen de condities, toonde een Kruskal-Wallis geen significant effect van

condities op het aantal representerende gebaren tijdens spraakverstoringen van $H(2) = .155, p = .925$. De Games-Howell post-hoc analyse (Tabel 4) toonde geen significante verschillen tussen de condities L1-L1 vs. L1-L2 ($M = .189, SE = .227, p = .686$), L1-L1 vs. L2-L2 ($M = .073, SE = .216, p = .939$) en L1-L2 vs. L2-L2 ($M = -.116, SE = .175, p = .788$). Dit betekent dat er geen verschil lijkt te zijn in het aantal representerende gebaren tijdens spraakverstoringen tussen de verschillende condities, wat niet in lijn is met de hypothese dat er tijdens spraakverstoringen in de L2-L2 conditie meer representerende gebaren worden gemaakt dan in de L1-L1 en L1-L2 conditie.

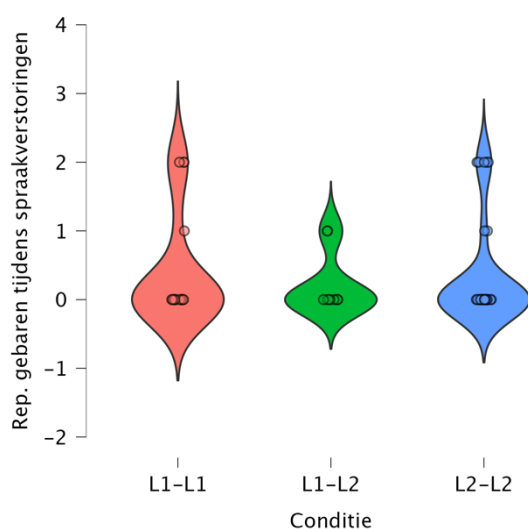
Tabel 3

Descriptives representerende gebaren tijdens spraakverstoringen per conditie

	Aantal representerende gebaren tijdens spraakverstoringen		
	L1-L1	L1-L2	L2-L2
N	18	10	38
Mediaan	0	0	0
IQR	0	0	0
Shapiro-Wilk	0.533	0.509	0.482
P-waarde van Shapiro-Wilk	< .001	< .001	< .001
Minimum	0	0	0
Maximum	2	1	2

Figuur 2

Vioolplot representerende gebaren tijdens spraakverstoringen per conditie



Tabel 4*Games-Howell Multiple Comparisons representerende gebaren tijdens spraakverstoringen*

Vergelijking	Vershil in gemiddelde	SE	t	df	P _{Tukey}
L1-L1 - L1-L2	0.189	0.227	0.833	26.000	0.686
L1-L1 - L2-L2	0.073	0.216	0.339	30.552	0.939
L1-L2 - L2-L2	-0.116	0.175	-0.661	23.816	0.788

Note. Resultaten zijn gebaseerd op niet-gecorrigeerde gemiddelden.

3.3 Aanvullende analyse

Uit de algemene analyse bleek dat participanten in de L1-L1 conditie niet significant minder representerende gebaren maakten dan participanten in de L1-L2 en L2-L2 conditie. Dit gold voor zowel voor representerende gebaren in het algemeen als enkel tijdens spraakverstoringen. Om deze bevinding nader te onderzoeken, is er eerst gekeken naar het algemene gebruik van representerende gebaren tijdens spraakverstoringen. In Tabel 5 is te zien dat het merendeel van de participanten – ongeacht conditie – geen enkel representerend gebaar maakte tijdens spraakverstoringen. In elke conditie maakten meer dan 75% van de participanten geen enkel representerend gebaar tijdens spraakverstoringen (zie Tabel 5 en Tabel 6).

Om te toetsen of er ondanks deze verdeling toch een significant verschil was tussen de condities in het gebruik van representerende gebaren (wel of geen gebaren) tijdens spraakverstoringen, werd er een Chi-kwadraat test uitgevoerd. Deze toonde geen significant resultaat ($\chi^2(2) = .112$, $p = .945$). De nominale contingentie coëfficiënt werd berekend om de sterkte van de associatie tussen de condities en het gebruik van representerende gebaren tijdens spraakverstoringen te meten. Die toonde een zwak effect van $C = .041$, wat betekent dat er geen sterke relatie lijkt te zijn tussen de condities en het gebruik van representerende gebaren tijdens spraakverstoringen.

Tabel 5*Frequenties voor aantal representerende gebaren tijdens spraakverstoringen*

Conditie	Aantal representerende gebaren tijdens spraakverstoringen	Frequentie	Percentage
L1-L1	0	14	77.778
	1	1	5.556
	2	3	16.667
	Totaal	18	100
L1-L2	0	8	80.000
	1	2	20.000
	Totaal	10	100
L2-L2	0	31	81.579
	1	2	5.263
	2	5	13.158
	Totaal	38	100

Tabel 6*Frequentietabel van representerende gebaren tijdens spraakverstoringen*

		Categorisatie van representerende gebaren tijdens spraakverstoringen	
Conditie		Geen gebaren	Wel gebaren
L1-L1	Aantal	14	4
	% in de rij	77.778 %	22.222 %
	% in de kolom	26.415 %	30.769 %
L1-L2	Aantal	8	2
	% in de rij	80.000 %	20.000 %
	% in de kolom	15.094 %	15.385 %
L2-L2	Aantal	31	7
	% in de rij	81.579 %	18.421 %
	% in de kolom	58.491 %	53.846 %

Omdat representerende gebaren tijdens spraakverstoringen zelden voorkwamen, is er gekeken naar een algemener patroon: maken sprekers überhaupt meer representerende gebaren tijdens spraakverstoringen dan tijdens vloeiende spraak, ongeacht de conditie? Hierbij werd er verwacht dat er juist tijdens spraak meer representerende gebaren zouden worden gemaakt.

In Tabel 7 is te zien dat er vooral tijdens spraak representerende gebaren gemaakt lijken te worden ($Mdn = 8$, $IQR = 15.500$) in tegenstelling tot representerende gebaren tijdens

spraakverstoringen ($Mdn = 0$, $IQR = 0$). Met een Shapiro-Wilk test is er gecontroleerd of de verdeling normaal verdeeld was. Deze toonde een significant resultaat voor zowel het aantal representerende gebaren tijdens spraak ($W = .801$, $p < .001$) als tijdens spraakverstoringen ($W = .500$, $p < .001$), wat duidt op een afwijking van normaliteit (zie Tabel 7). Daarom werd er een Wilcoxon signed-ranked test gedaan, die een significant verschil tussen de condities toonde ($Z = 6.791$, $p < .001$). Dit betekent dat er significant meer representerende gebaren tijdens spraak werden gemaakt dan tijdens spraakverstoringen.

Tabel 7

Descriptives aantal representerende gebaren tijdens spraak en spraakverstoringen

	Aantal representerende gebaren tijdens spraak	Aantal representerende gebaren tijdens spraakverstoringen
N	66	66
Mediaan	8	0
IQR	15.500	0
Shapiro-Wilk	0.801	0.500
P-waarde van Shapiro-Wilk	< .001	< .001
Minimum	0	0
Maximum	60	2

4.0 Discussie

De huidige studie is gedaan om meer kennis op te doen over de relatie tussen representerende gebaren en spraakverstoringen. Er is onderzocht in welke mate het maken van representerende gebaren samenhangt met spraakverstoringen door de condities L1-L1, L1-L2 en L2-L2 te vergelijken.

4.1 Interpretatie met literatuur

We vonden geen verschillen tussen de L1-L1, L1-L2 en L2-L2 condities in het aantal representerende gebaren. Dit is niet in lijn met onderzoek van Hadar et al. (2001) en Arslan et al. (2024), waaruit bleek dat mensen wel degelijk meer gebaren gebruikten in hun L2 dan in hun L1. Op basis daarvan werd er verwacht dat er in de L1-L2 en L2-L2 condities meer

representerende gebaren zouden worden gemaakt dan in de L1-L1 conditie, maar dit bleek niet het geval. Holler & Beattie (2003) vonden dat sprekers representerende gebaren gebruikten om verbale ambiguïteit te verhelderen. Met name het bewustzijn van een mogelijk communicatieprobleem bleek bepalend voor het gebruik van representerende gebaren (Holler & Beattie, 2003). Hoewel er in onze studie werd gefocust op spraakverstoringen in plaats van verbale ambiguïteit, zou dit kunnen verklaren waarom we geen significante resultaten vonden in de algemene analyse: er werd niet gecontroleerd op taalvaardigheid. Een ander verschil met ons onderzoek is dat in eerdere studies participanten werden vergeleken op basis van de taal waarin zij spraken (L1 of L2), ongeacht hun gesprekspartner. In het huidige onderzoek zijn participanten vergeleken op basis van hun conditie, waarbij de combinaties van sprekers – L1-L1, L1-L2 of L2-L2 – werden vergeleken. Dit verschil in opzet kan verklaren waarom onze resultaten afwijken van eerder onderzoek, omdat in ons ontwerp ook de invloed van de gesprekspartner is meegenomen.

We vonden ook geen verschillen tussen de condities in aantal representerende gebaren tijdens spraakverstoringen. Dit is in strijd met *lexical retrieval hypothesis*, die stelt dat gebaren de sensomotorische kenmerken van op te halen woorden activeren, waardoor deze gemakkelijker worden opgehaald en geproduceerd (Krauss, 1998; Arslan et al., 2024). Dit zou betekenen dat er juist in situaties waarin de taalproductie moeilijker is – zoals wanneer iemand in een L2 spreekt – er meer representerende gebaren worden gemaakt ter ondersteuning van de woordophaling ten opzichte van iemand die een L1 spreekt. Het feit dat er geen verschil is gevonden tussen de condities, suggereert dat gebaren niet ondersteunend waren voor de taalproductie tijdens spraakverstoringen. Ook is dit in strijd met Arslan et al. (2024), die aantoonde dat gebaren tijdens spraakverstoringen worden gebruikt als sprekers in staat zijn hun eigen fouten te herkennen. Dit zou betekenen dat de taalvaardigheid van sprekers invloed heeft op het gebruik van gebaren tijdens spraakverstoringen. Oftewel, hoe beter iemand de taal

beheerst, hoe groter de kans dat iemand zich bewust is van een fout of spraakverstoring en daarom een gebaar gebruikt om de communicatie te verbeteren. Vanuit dit perspectief werd er verwacht dat in de L1-L2 en L2-L2 conditie meer representerende gebaren worden gebruikt in vergelijking met de L1-L1 conditie ter compensatie voor communicatieproblemen. Een verklaring voor het verschil met eerdere bevindingen is dat de taalvaardigheid van participanten – die volgens eerder onderzoek invloed heeft op het gebruik van representerende gebaren tijdens spraakverstoringen (Arslan et al., 2024) – in de huidige analyse niet is meegenomen. Een medestudent heeft taalvaardigheid wél meegenomen in haar analyse, waaruit bleek dat participanten met een hoge taalvaardigheid meer iconische gebaren – een vorm van representerende gebaren – gebruikten tijdens spraakverstoringen dan participanten met een lage taalvaardigheid. Deze resultaten lijken het onderzoek van Arslan et al. (2024) te ondersteunen, wat suggereert dat taalvaardigheid mogelijk een moderator is voor representerende gebaren tijdens spraakverstoringen.

De resultaten van de aanvullende analyse – waarbij er tijdens vloeiende spraak significant meer representerende gebaren werden gemaakt dan tijdens spraakverstoringen – zijn in lijn met het onderzoek van Graziano en Gullberg (2018). Zij vonden dat gebaren stopten wanneer spraak stopte en dat gebaren met name werden gebruikt tijdens vloeiende spraak in plaats van spraakverstoringen. Deze bevindingen staan haaks op onderzoek van Arslan et al. (2024), Hadar et al. (2001) en de *lexical retrieval hypothesis* (Krauss, 1998; Arslan et al., 2024). Onze resultaten ondersteunen het onderzoek van Graziano en Gullberg (2018): gebaren lijken inderdaad te stoppen als spraak stopt. Er is in de aanvullende analyse echter niet gecontroleerd voor een eventueel verschil tussen L1- en L2-sprekers. Vervolgonderzoek is daarom nodig om deze resultaten te valideren.

4.2 Sterke punten

Aan het experiment namen 66 participanten deel. Deze steekproefgrootte is groter dan of vergelijkbaar met andere studies binnen dit onderzoeksveld (Arslan et al., 2024; Bergmann et al., 2015; Graziano & Gullberg, 2018), wat de resultaten goed vergelijkbaar maakt met eerder onderzoek. Daarnaast werd de betrouwbaarheid van het coderen vergroot door te coderen volgens de richtlijnen van de MultiModal MultiDimensional Labeling Sytem (Tütüncübasi et al., 2023). De onderzoekers volgden hier een online training om bekend te raken met het coderen, zodat er zo betrouwbaar mogelijk kon worden gecodeerd. Ook droeg de duidelijke onderzoeksstructuur bij aan een betrouwbare methodologie, wat de reproduceerbaarheid van het onderzoek vergroot.

4.3 Limitaties

Aangezien de participanten in Nederland studerende studenten waren, is de sample uit een WEIRD-populatie (Jones, 2024) getrokken. Dit maakt het aannemelijk dat het onderzoek niet generaliseerbaar is naar andere culturen. Uit eerder onderzoek blijkt namelijk dat er vier factoren invloed hebben op de culturele verschillen in het gebruik van gebaren – namelijk cultuurspecifieke conventies over de betekenis van gebaren, verschillen in ruimtelijke cognitie, linguïstische verschillen en het pragmatische gebruik van gebaren (Kita, 2009). In het huidige onderzoek is er geen rekening gehouden met deze factoren. Daarnaast namen de participanten in tweetallen deel aan het onderzoek, waardoor de observaties niet volledig onafhankelijk waren. Hierdoor kunnen de participanten elkaar – en daarmee de onderzoeksresultaten – hebben beïnvloed. Bovendien vormt de ecologische validiteit een beperking, aangezien het experiment in een laboratoriumsetting is uitgevoerd. Ook werden spraak en gebaren handmatig gecodeerd door de onderzoekers, die voorafgaand geen tot weinig ervaring hadden met coderen. Dit maakt het aannemelijk dat er – ondanks het gebruik van het M3D coderingssysteem – fouten zijn gemaakt in bijvoorbeeld het coderen van het type gebaren. Dit kan hebben gezorgd voor ruis.

Hieraan gerelateerd is de categorisering van gebaren subjectief. Hoewel er wordt gewerkt aan de standaardisering van gebaren- en spraakcodering, is dit momenteel nog niet goed genoeg gestandaardiseerd. Zo wordt er aangegeven in de M3D training dat gebaren op meerdere manieren te interpreteren zijn, wat de subjectiviteit bevestigt (Rohrer et al., 2023).

4.4 Vervolgonderzoek en implicaties

Er zijn verschillende onduidelijkheden, die in een vervolgonderzoek opgehelderd dienen te worden. Allereerst zijn in het huidige onderzoek de verschillende condities vergeleken in plaats van de individuele participanten. Zo is er bij participanten in de L1-L2 conditie geen onderscheid gemaakt tussen participanten die hun L1 of L2 spraken. Dit kan van invloed zijn geweest op de resultaten. Daarom wordt er aanbevolen om in een vervolgonderzoek individuele participanten in plaats van condities te vergelijken. Daarnaast bleek uit eerder onderzoek dat er mogelijk enkel gebaren worden gebruikt tijdens spraakverstoringen als sprekers in staat zijn hun fouten te herkennen (Arslan et al., 2024). Ook lijken de bevindingen van een medestudent – die controleerde op taalvaardigheid – dit te ondersteunen. Dit dient met een vervolgonderzoek te worden opgehelderd door het aantal representerende gebaren tijdens spraakverstoringen van L1- en L2-sprekers te vergelijken, waarbij er wordt gecontroleerd op taalvaardigheid. Ook dient er onderzocht te worden of het inderdaad zo is dat gebaren stoppen wanneer spraak stopt, wat eerder werd aangetoond in onderzoek van Graziano en Gullberg (2018) en Kisa et al. (2021). De aanvullende analyse lijkt deze resultaten te bevestigen, maar om dit te kunnen stellen is verder onderzoek noodzakelijk. Tot slot is er niet onderzocht of culturele verschillen invloed hebben op de frequentie van gebaren. Door in een vervolgonderzoek verschillende culturen met elkaar te vergelijken, kan op deze vraag antwoord worden gegeven. Ook is het interessant om hierbij de taalvaardigheid van sprekers mee te nemen, aangezien blijkt uit onderzoek van Arslan et al. (2024) en een medestudent dat dit invloed kan hebben op het gebruik van gebaren tijdens spraakverstoringen.

4.5 Conclusie

In deze studie is onderzocht in welke mate het maken van representerende gebaren samenhangt met spraakverstoringen door de condities L1-L1, L1-L2 en L2-L2 te vergelijken. Op basis van het huidige onderzoek lijkt er geen samenhang te zijn tussen het aantal van representerende gebaren en spraakverstoringen in de condities L1-L1, L1-L2 en L2-L2. Er werd geen significant verschil gevonden voor het aantal representerende gebaren tussen de condities L1-L1, L1-L2 en L2-L2. Ook werd er geen significant verschil gevonden in het aantal representerende gebaren tijdens spraakverstoringen tussen de verschillende condities. Dit is in strijd met eerder onderzoek van Hadar et al. (2001), Arslan et al. (2024) en de *lexical retrieval hypothesis* (Krauss, 1998; Arslan et al., 2024). Uit een aanvullende analyse bleek echter dat er tijdens vloeiende spraak significant meer gebaren werden gemaakt dan tijdens spraakverstoringen, wat onderzoek van Graziano en Gullberg (2018) ondersteund. Hierbij is echter niet gecontroleerd op verschillen tussen L1- en L2-sprekers. Ook is de taalvaardigheid van participanten niet meegenomen. Vervolgonderzoek dient uit te wijzen of gebaren inderdaad stoppen tijdens spraakverstoringen en of er hierbij verschillen zijn tussen L1- en L2-sprekers, terwijl er wordt gecontroleerd voor de taalvaardigheid.

Referenties

- Arslan, B., Avcı, C., Yılmaztekin, A., & Göksun, T. (2024). Do bilingual adults gesture when they are disfluent?: Understanding gesture-speech interaction across first and second languages. *Language Cognition And Neuroscience*, 39(5), 571–583. <https://doi.org/10.1080/23273798.2024.2345306>
- Audacity Team. (2014). *Audacity(R): Free Audio Editor and Recorder* (3.7.3) [Software]. <http://audacity.sourceforge.net/>
- Bergmann, C., Sprenger, S. A., & Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychologica*, 161, 25-35. <https://doi.org/10.1016/j.actpsy.2015.07.015>
- Coleman, J., Grotjahn, R., Raatz, U., & Wockenfuß, V. (2017a). *Gap-fill / Cloze test / Fill in the blank from “Das TESTATT-Projekt: Entwicklung von C-Tests zur Evaluation des Fremdsprachenlernerfolgs”* [Text/Language test]. IRIS Database, University of York, UK. <https://www.iris-database.org/details/7ZlyH-wLAPu>
- Coleman, J., Grotjahn, R., Raatz, U., & Wockenfuß, V. (2017b). *Gap-fill / Cloze test / Fill in the blank from “Das TESTATT-Projekt: Entwicklung von C-Tests zur Evaluation des Fremdsprachenlernerfolgs”* [Text/Language test]. IRIS Database, University of York, UK. <https://www.iris-database.org/details/M3rMX-RKx9r>
- De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency. *IRAL - International Review Of Applied Linguistics in Language Teaching*, 54(2). <https://doi.org/10.1515/iral-2016-9993>
- ELAN (Version 6.9). (2024). [Software]. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. <https://archive.mpi.nl/tla/elan>
- Freeborn, L., & Andringa, S. (2023). Development of a C-test to measure Dutch language proficiency. *OSF*. <https://doi.org/10.17605/OSF.IO/VNTG3>

- Gilmore, A. (2013). *C-test. Gap-fill / Cloze test / Fill in the blank from “I prefer not text”*: *Developing Japanese learners' communicative competence with authentic materials* [Text/Language test]. IRIS Database, University of York, UK. <https://www.iris-database.org/details/GXgIl-kFpif>
- Graziano, M., & Gullberg, M. (2018). When Speech Stops, Gesture Stops: Evidence From Developmental and Crosslinguistic Comparisons. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00879>
- Gullberg, M. (2022). The Relationship Between Gestures and Speaking in L2 Learning. In *Routledge eBooks* (pp. 386–398). <https://doi.org/10.4324/9781003022497-33>
- Gullberg, M., De Bot, K., & Volterra, V. (2008). Gestures and some key issues in the study of language development. *Gesture*, 8(2), 149–179. <https://doi.org/10.1075/gest.8.2.03gul>
- Hadar, U., Dar, R., & Teitelman, A. (2002). Gesture during speech in first and second language. *Gesture*, 1(2), 151–165. <https://doi.org/10.1075/gest.1.2.04had>
- Holler, J., & Beattie, G. (2003). Pragmatic aspects of representational gestures. *Gesture*, 3(2), 127–154. <https://doi.org/10.1075/gest.3.2.02hol>
- JASP. (2025). [Software]. In *JASP - Free and User-Friendly Statistical Software* (0.19.3). <https://jasp-stats.org/>
- Jones, D. (2024). A WEIRD view of human nature skews psychologists' studies. In *American Psychological Association eBooks*(pp. 163–165). <https://doi.org/10.1037/0000409-006>
- Kendon, A. (2004). *Gesture Visible Action as Utterance*. <https://doi.org/10.1017/cbo9780511807572>
- Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language And Cognitive Processes*, 24(2), 145–167. <https://doi.org/10.1080/0169090802586188>
- Krauss, R.M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, 7(2), 54. <https://doi.org/10.1111/1467-8721.ep13175642>

- Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15(1), 19–44. <https://doi.org/10.1080/00437956.1959.11659682>
- Mayberry, R. I., & Jaques, J. (2000). Gesture production during stuttered speech: insights into the nature of gesture–speech integration. In *Cambridge University Press eBooks* (pp. 199–214). <https://doi.org/10.1017/cbo9780511620850.013>
- McNeill, D. (1992). Hand and mind: what gestures reveal about thought. *Choice Reviews Online*, 30(05), 30–2492. <https://doi.org/10.5860/choice.30-2492>
- Nicoladis, E., Nagpal, J., Marentette, P., & Hauer, B. (2018). Gesture frequency is linked to story-telling style: evidence from bilinguals. *Language And Cognition*, 10(4), 641–664. <https://doi.org/10.1017/langcog.2018.25>
- Rohrer, P. L., Tütüncübasi, U., Vilà-Giménez, I., Florit-Pons, J., Esteve-Gibert, N., Ren-Mitchell, A., Shattuck-Hufnagel, S. & Prieto, P. (2023). The MultiModal MultiDimensional (M3D) labeling system. <https://doi.org/10.17605/osf.io/ankdx>
- Tütüncübasi, U., Ren-Mitchell, A., Shattuck-Hufnagel, S., Rohrer, P. L. & Prieto, P. (2023). The M3D Training Program. <https://m3d.upf.edu>
- Tree, J. E. (1995). The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal Of Memory And Language*, 34(6), 709–738. <https://doi.org/10.1006/jmla.1995.1032>
- Wagner, P., Malisz, Z., & Kopp, S. (2013). Gesture and speech in interaction: A overview. *Speech Communication*, 57, 209–232. <https://doi.org/10.1016/j.specom.2013.09.008>
- Qualtrics XM - *Experience Management Software*. (2025, April 15). Qualtrics. <https://www.qualtrics.com/>

Bijlage A

Stellingen

- 1) Which came first, the chicken or the egg?
- 2) Would you rather have a job you dislike with a good salary or have a job you like with a bad salary?
- 3) Does technology improve human life overall, or does it cause more harm than good in the long run?
- 4) Would you rather be very funny or very intelligent?
- 5) Is it better to know the truth, even if it makes you unhappy, or to live blissfully ignorant of unpleasant realities?
- 6) Would you rather only write papers/assignments or only make exams for school?
- 7) If you replace all parts of a ship over time, is it still the same ship? If not, at what point did it cease to be the original?
- 8) Would you rather have no summers or no winters for the rest of your life?
- 9) Is progress driven more by original innovations or by improvements on existing ideas?
- 10) Do humans truly have free will to make choices, or are all events and decisions predetermined by prior causes?

Bijlage B**Figuur 1***Q-Q plots representerende gebaren per conditie*