

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/150789>

Please be advised that this information was generated on 2019-06-16 and may be subject to change.

# Spraakherkenning van kinderstemmen

*In een nummer over kinderen en TST mag natuurlijk niet een stuk ontbreken over het herkennen van de spraak van kinderen. Is dat iets apart dan? Kinderen praten toch gewoon als volwassenen alleen dan met hogere stemmen, net als vrouwen, maar dan een beetje anders?*

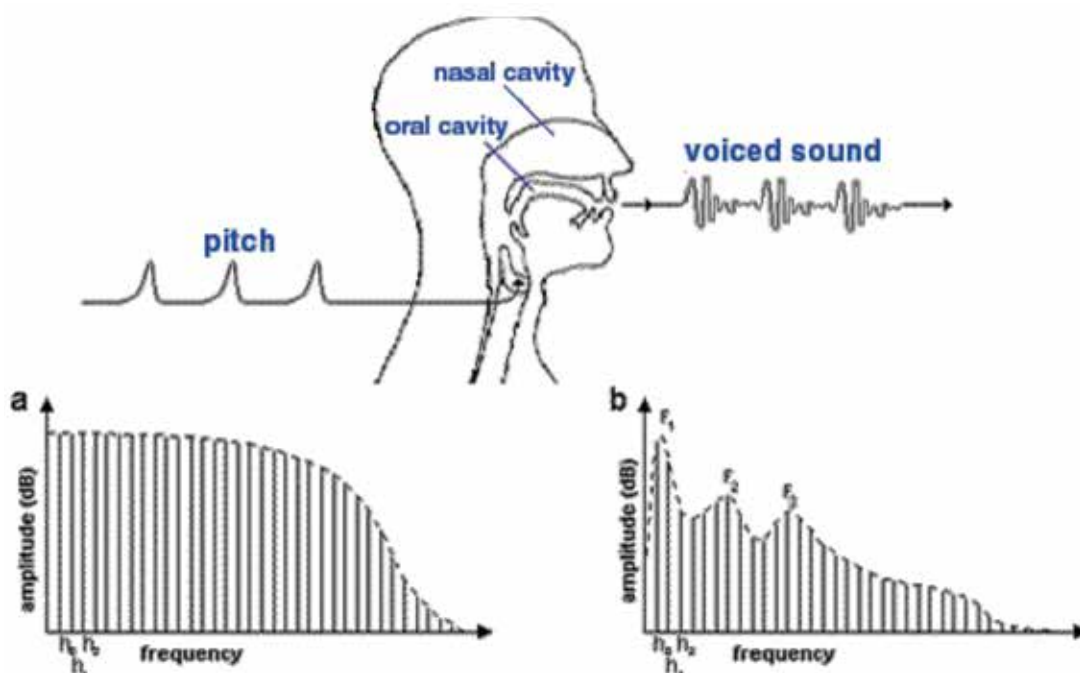
**Arjan van Hessen**  
**Telecats en**  
**Universiteit**  
**Utrecht**  
**Louis ten Bosch**  
**CLST en**  
**Radboud**  
**Univesiteit**  
**Nijmegen**

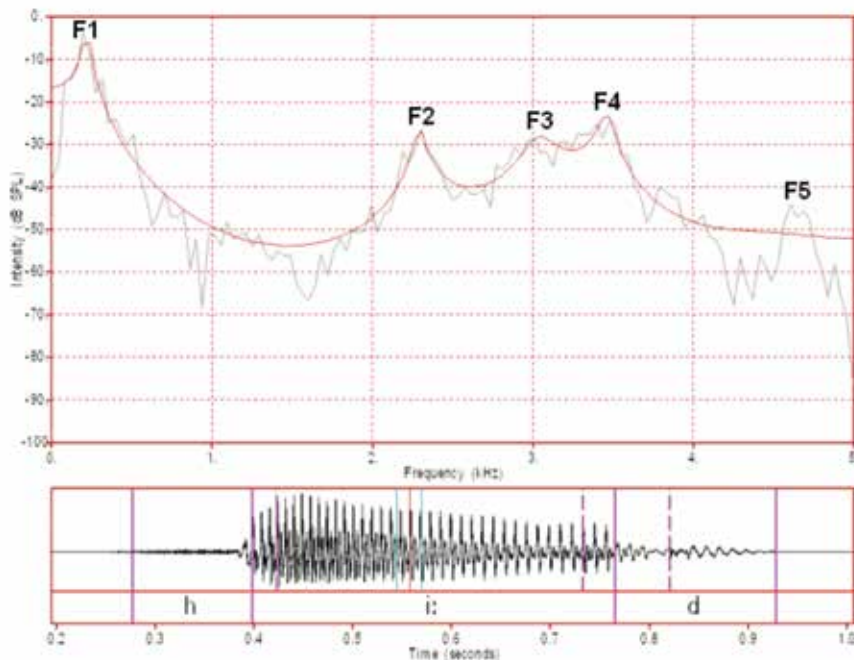
Deels is dat natuurlijk zo, maar niet helemaal. Om duidelijk te maken waarom het herkennen van kinderspraak zo lastig is, geven we eerst een kort overzicht van de manier waarop het bij 'gewone mensen' gaat. Dit leidt uiteindelijk tot de conclusie dat computers apart getraind moeten worden voor het herkennen van kinderspraak.

## Spraak

Mensen praten door lucht door hun strottenhoofd te blazen. Door je stembanden mee te laten trillen krijg je een klinker, door dat niet te doen, krijg je een medeklinker (medeklinkers zijn iets complexer, maar dat doet er hier niet toe). De stembanden die trillen, zorgen voor een 'grondtoon' van  $N$  Hertz: dat houdt in dat de stembanden  $N$  keer per seconde open en dicht gaan. Bij mannen ligt de  $N$  tussen de 80 en 150, bij vrouwen tussen de 150 en 250 Hz. Net als bij een muziekinstrument, krijg je dan in principe trillingen van  $N$ ,  $2N$ ,  $3N$ ,  $4N$  tot  $\infty N$ . Al die trillingen gaan langs je stembanden omhoog, door je keel en mond en dan naar buiten. Wanneer alle trillingen even luid de mond verlaten, dan is het resulterende signaal een zogeheten spike.

Maar dat gebeurt niet. De trillingen komen in je keel en mond tegen je wangen, je tanden, tong en gehemelte aan waardoor er reflecties ontstaan. Sommige trillingen zullen dan juist versterkt worden, andere juist uitgedoofd. Het geluidssignaal dat dan naar buiten komt, is het resultaat van al die optredende reflecties in je mond. Door de stand van je mond te veranderen, verander je de reflectiepatronen en dus de klank die naar buiten komt. Probeer maar eens een luide /A/ te zingen en steek dan je wijsvinger in je mond zonder je mond of tong aan te raken. Hoe verder je je vinger je mond in duwt, hoe meer de /A/ verandert. Dat komt doordat de trillingen in je mond tegen je vinger aan komen waardoor de reflecties veranderen. Precies op deze manier maken wij mensen spraak (klinkers). Nu zijn de reflecties in je mond afhankelijk van de vorm en dus van de grootte van je mond en keelkanaal. Dit is vergelijkbaar met een muziekinstrument. Een viool klinkt heel anders dan een cello, ook al gebruik je dezelfde snaar en is de toonhoogte gelijk. En probeer maar eens een paar opgerolde sokken in de klankkast van een viool te stoppen en luister hoe het dan klinkt!





### Formanten

Volwassen mannen hebben in de regel vijf duidelijke reflectiepunten in hun mond-keelkanaal waardoor je vijf zogeheten formanten in de spectrale weergave van hun spraak ziet. Door de mond-keel te vervormen, komen die reflectiepunten op andere plekken te liggen, maar het blijven er (meestal) vijf. Door de positie van je mond te veranderen, veranderen de reflectiepunten en hoor je een andere klank (=klinker). Vrouwen hebben kortere stembanden en een kleiner mond-keelkanaal waardoor de spraak van vrouwen iets anders klinkt (en niet alleen hoger). Dit merk je bij transseksuelen. De stembanden worden naar gelang de geslachtsverandering ingekort/verlengd maar toch hoor je vaak bij transseksuelen 'dat er iets aan de stem niet klopt'. Dit is juist want het mond-keelkanaal kun je moeilijk aanpassen (zie: [www.kennislink.nl/publicaties/genderradar-gaat-af-op-spreekstijl](http://www.kennislink.nl/publicaties/genderradar-gaat-af-op-spreekstijl)).

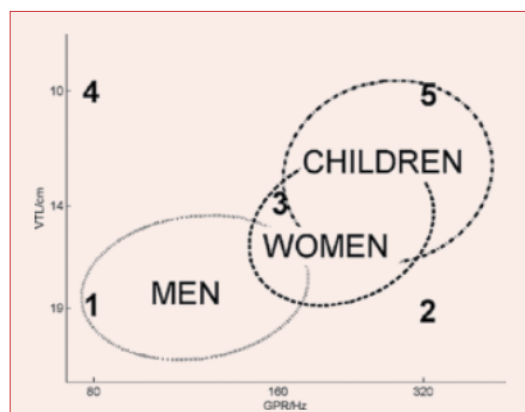
### Kinderen

Kinderen zijn in de regel veel kleiner dan volwassenen, ze hebben dus kortere stembanden waardoor de toonhoogte hoger is: 260 Hz voor kleine kinderen terwijl die op ongeveer 90 Hz ligt bij grote volwassen mannen. Maar ook het mond-keelkanaal van kinderen is veel kleiner waardoor de reflecties weer anders zijn. De lengte van het mond-keelkanaal varieert van 9 centimeter bij jonge kinderen tot 17 centimeter bij grote (jong) volwassenen. Anders dan bij de stembanden is de lengte van het mond-keelkanaal gekoppeld aan de lengte van de persoon en niet aan het geslacht. Toch kunnen we met kinderen praten, dus blijktbaar kunnen zij hun mond-keelkanaal zo vormen, dat de spraak die daar uitkomt voor

ons verstaanbaar is en kunnen wij onze oren 'aanpassen' aan de spraak van kinderen.

Maar er is meer. Kinderen hebben gemiddeld een minder goed ontwikkeld taalgevoel dan volwassenen en dus is hun woordgebruik anders. Ze gebruiken iets andere woorden en knopen woorden soms op een andere manier aan elkaar.

Kortom: kinderen spreken anders en hebben een andere stem dan volwassenen. De moderne spraakherkenners zijn getraind op de spraak van volwassenen en daarom hebben ze moeite met de spraak van kinderen. We kunnen daar iets aan doen door computers te trainen met kinderspraak, maar dan moeten we wel eerst voldoende data hebben en dat kost geld.



Ellipses showing the distribution of vocal tract lengths and glottal pulse rates for men, women and children, based on the data of Peterson and Barney (1952). The ellipses encompass 96 % of the data for each group. The numbered points show the reference speakers used in the experiment.