

Stemmen meten met Praat

Paul Boersma, Universiteit van Amsterdam, 23 december 2004

Abstract

Dit artikel beschrijft hoe het computerprogramma Praat gebruikt kan worden voor het meten van verscheidene soorten afwijkingen in de periodiciteit van een stemgeluid, te weten stemhaperingen, jitter, shimmer en 'ruizigheid'. De metingen in Praat blijken over een groter gebied betrouwbaar, en minder gevoelig voor toegevoegde ruis, dan de metingen in het programma MDVP. Pas door deze toegenomen nauwkeurigheid kunnen jitter- en shimmermetingen correct gerelateerd worden aan specifieke onregelmatigheden van de stem.

This article describes how the computer program Praat can be used to measure several kinds of deviations from the periodicity of the human voice, to wit voice breaks, jitter, shimmer, and 'noisiness'. Measurements with Praat turn out to be reliable over a larger range, and less sensitive to additive noise, than measurements with the program MDVP. Only with this improved accuracy can jitter and shimmer measurements be interpreted as reflecting specific irregularities in the voice.

Dit artikel beschrijft hoe U een aantal soorten afwijkingen in de periodiciteit van het stemgeluid (namelijk stemhaperingen, jitter, shimmer en 'ruizigheid') kunt meten met het computerprogramma Praat, en hoe betrouwbaar deze metingen zijn. Voor drie van deze soorten afwijkingen (namelijk jitter, shimmer en 'ruizigheid') geldt dat er een groot aantal definities zijn te vinden in de literatuur, die ook de vaak complexe relaties beschrijft tussen deze periodiciteitsafwijkingen en specifieke pathologieën (zoals tumoren en verlammingen) of stemkenmerken (zoals heesheid en schorheid). Zie bijvoorbeeld Lieberman (1963), Wendahl (1966), Koike (1969), Iwata & Von Leden (1970), Klingholz & Martin (1983), Hillenbrand (1987). Een direct klinisch belang van deze metingen is dat ze objectieve maten vormen voor stemproblemen die vaak het eerste teken zijn van een beginnende pathologie. Verder vormen ze bijvoorbeeld objectieve maten voor de kwaliteit van de stem in een revalidatietraject, bijvoorbeeld na radiotherapie (Verdonck-De Leeuw 1998) of laryngectomie (Van As 2001).

1. Een perfect periodiek stemgeluid

Figuur 1 toont bijna zes *periodes* van de golfvorm van een perfect periodiek geluid met een grondfrequentie (toonhoogte) van 117 Hertz. Dit geluid klinkt als de klinker [a], dus het zou gemaakt kunnen zijn door stembanden die zich perfect regelmatig 117 keer per seconde sluiten en openen, telkens met precies dezelfde energie en zonder toegevoegde ruis, terwijl bovendien de tong, de lippen, de kaak en de keelholte de hele tijd in dezelfde stand en vorm worden gehouden.

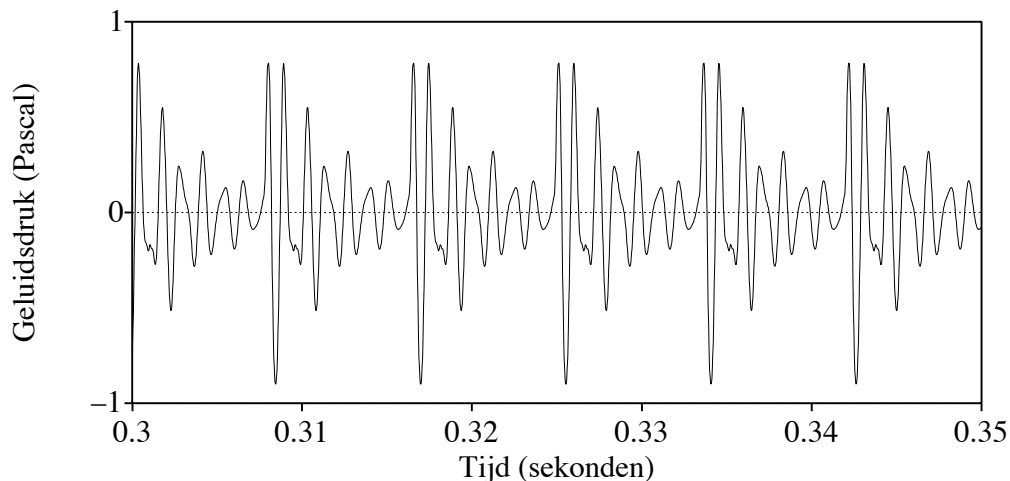


Fig. 1. Een perfect periodiek stemgeluid.

Zo'n perfect periodiek stemgeluid kan geen mens produceren, maar gezonde stemmen komen wel in de buurt. Patiënten met larynxproblemen kunnen daarentegen grote afwijkingen vertonen. Dit artikel beschrijft hoe een aantal van die afwijkingen kunnen worden gemeten met het computerprogramma Praat, dat is ontwikkeld door Paul Boersma en David Weenink van het Instituut voor Fonetische Wetenschappen van de Universiteit van Amsterdam. Nadat ik uitleg hoe u in Praat een stemrapport verkrijgt (§2), beschrijf ik de meting van de volgende vier soorten afwijkingen van de periodiciteit: stemhaperingen (§3), jitter (§4), shimmer (§5) en 'ruizigheid' (§6). De kwaliteit van Praat voor deze metingen wordt vergeleken met de kwaliteit van metingen onafhankelijk uitgevoerd door het computerprogramma dat het meest gebruikt wordt in de klinische praktijk, namelijk Multi Dimensional Voice Program™ (MDVP) van Kay Elemetrics®.

2. Een stemrapport in Praat

Het programma Praat is beschikbaar voor alle veel gebruikte computersystemen, waaronder Windows, Macintosh en Linux, en is vrij te downloaden via www.praat.org. Na de eenvoudige installatie van het programma (het bestaat uit één enkel bestand), leest u in Praat's **Intro** (onder het Help-menu) hoe u een geluid opneemt of een geluidsbestand inlaadt. Praat zal het geluid laten zien in een *geluidsvenster* (Figuur 2) waar in eerste instantie de golfvorm te zien is met daaronder drie analyses: een *spectrogram* (in grijsstinten), een *toonhoogtecontour* (een licht- en donkerblauwe curve, onderbroken waar het geluid stemloos is) en *formantcontouren* (rode stippen; het geluid in de Figuren 2, 3 en 5 is overigens een pathologisch stemgeluid beschikbaar gesteld door Maria Cristina Jackson Menaldi van Wayne University, Detroit). U kunt inzoomen, uitzoomen, selecteren (door met de muis te slepen) en scrollen, zodat u elk onderdeel van het geluid goed kunt bekijken en beluisteren. De formantcontouren kunt U indien gewenst uitschakelen (met **Show formants** in het Formant-menu) en het spectrogram misschien ook (met **Show spectrogram** in het Spectrum-menu). De toonhoogtecontour moet echter zichtbaar blijven want die is wezenlijk voor het verkrijgen van een stemrapport.

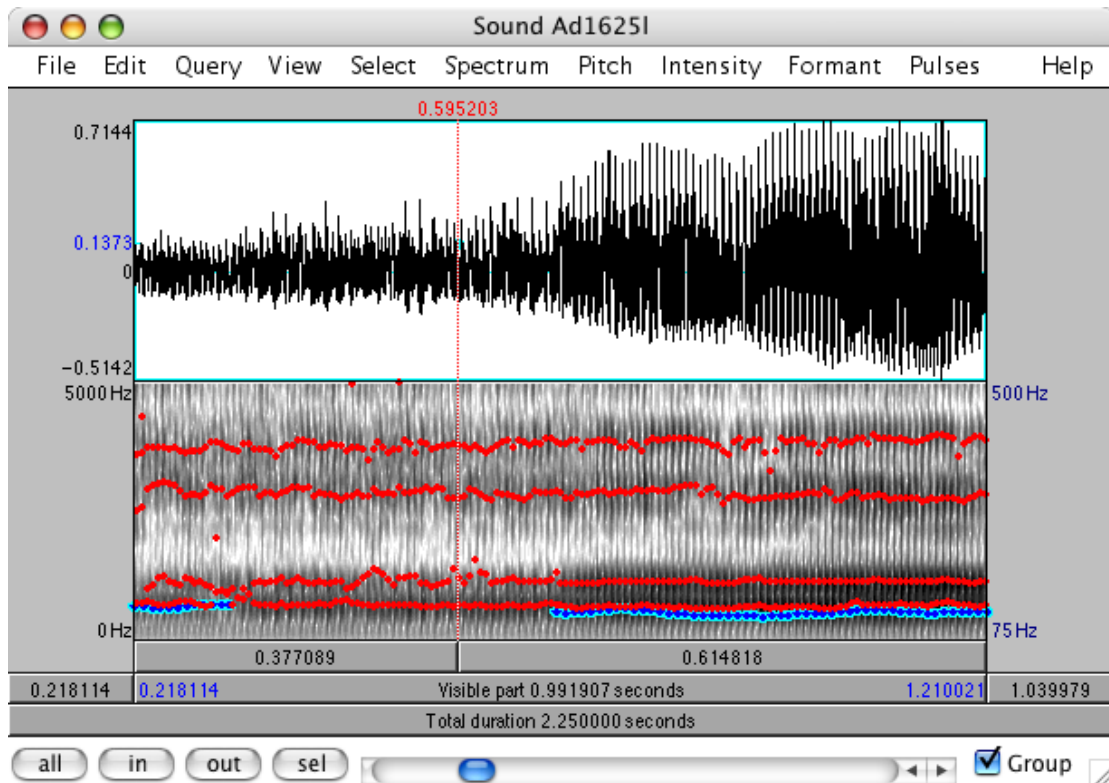


Fig. 2. Het geluidsvenster in Praat met standaardinstellingen.

De standaardinstellingen van Praat zijn optimaal voor linguïstisch georiënteerd spraakonderzoek. De toonhoogtecontour wordt bijvoorbeeld gemeten met een procedure die geoptimaliseerd is voor het onderzoek naar intonatie, waarbij de kortsturende variaties in de duur van de periodes (veroorzaakt door bv. stemzetting of naburige consonanten) genegeerd moeten kunnen worden. Voor onderzoek naar pathologische stemmen zijn deze microvariaties juist wel weer van groot belang, omdat ze daar zelfs kunnen voorkomen bij aangehouden klinkers en in dat geval een diagnostisch hulpmiddel kunnen zijn. Daarom is het hier van belang dat de toonhoogtemetingen worden gemeten met een procedure die de golfvormen van afzonderlijke opeenvolgende periodes met elkaar vergelijkt. Om de toonhoogtemetingen te optimaliseren voor stemanalyse gaat u naar **Pitch settings** (in het Pitch-menu) en kiest u **Optimize for voice analysis** in plaats van **Optimize for intonation** (als u dit vergeet krijgt u later in het stemrapport een waarschuwing dat sommige gegevens minder nauwkeurig kunnen zijn, zodat u de instelling alsnog kunt veranderen). De toonhoogtecontour wordt nu vier keer langzamer berekend dan voordien, omdat een grotere mate van detail moet worden onderzocht.

Om een stemrapport te verkrijgen moet niet alleen de toonhoogtecontour zichtbaar zijn, maar moeten ook de tijdstippen van de stembandpulsen zichtbaar worden gemaakt. Dat doet u met **Show pulses** in het Pulses-menu. De pulsen verschijnen als verticale blauwe lijnen door de golfvorm, zoals in Figuur 3. De genoemde twee veranderingen aan de instellingen van Praat (optimalisatie voor pathologie en zichtbaar maken van de stembandpulsen) hoeft u maar één keer te doen, want Praat onthoudt deze instellingen als u het geluidsvenster sluit en zelfs als u het programma verlaat.

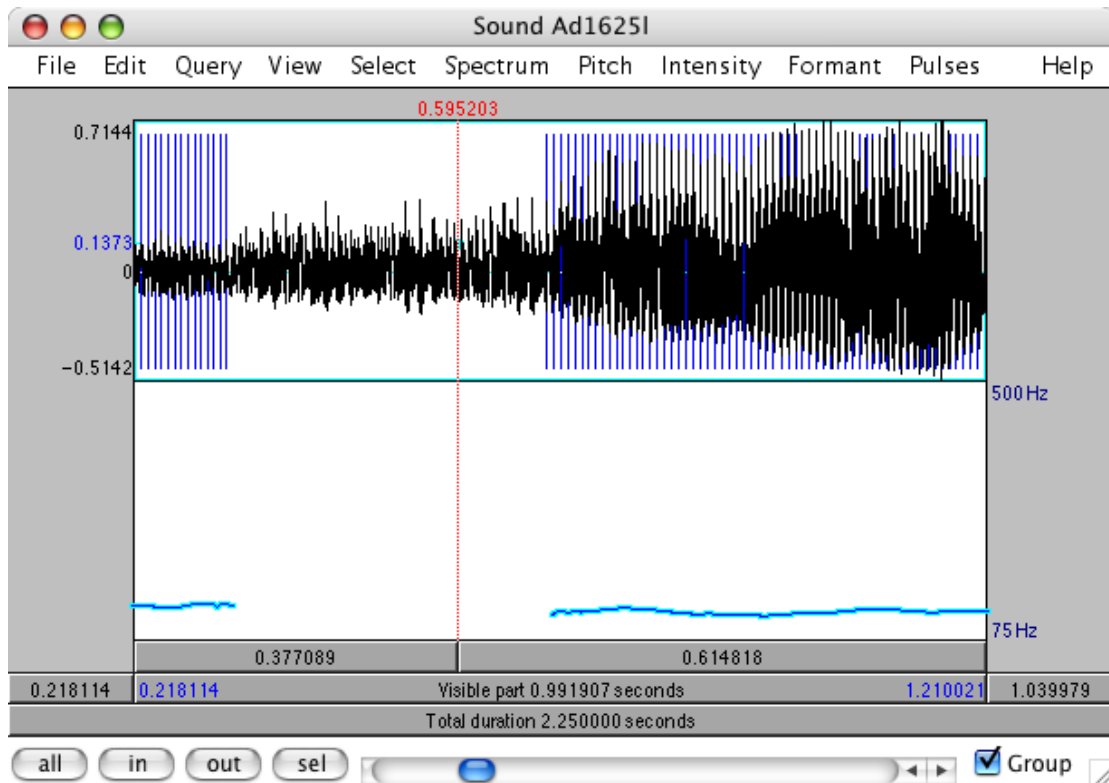


Fig. 3. Het geluidsvenster in Praat, geschikt gemaakt voor stemanalyse.

Om nu een stemrapport te krijgen selecteert u (door met de muis te slepen) het relevante gedeelte van het geluid (bijvoorbeeld het gedeelte dat een poging tot aangehouden klinker voorstelt) en kiest u **Voice report** in het Pulses-menu. Als u niets selecteert krijgt u een stemrapport over het zichtbare deel van het venster. Figuur 4 laat een voorbeeld zien. Het is het stemrapport over het zichtbare gedeelte van het geluid in Figuur 3. In Figuur 4 zien we eerst de informatie over begin- en eindtijd van de analyse (ook zichtbaar in Figuur 3), en vervolgens wat statistieken over de toonhoogtewaarden in de (onderbroken) toonhoogtecontour die zichtbaar is in Figuur 3: mediaan en gemiddelde toonhoogte over de 298 gemeten toonhoogtewaarden, de standaarddeviatie over deze 298 waarden, en de kleinste en grootste van deze 298 waarden. Vervolgens is er informatie over de stembandpulsen: hun aantal en de gemiddelde periode, d.w.z. de gemiddelde tijdsafstand tussen twee opeenvolgende pulsen (lange onderbrekingen worden niet meegeteld, zie §3).

```

-- Voice report for Sound Ad16251 --
Date: Fri Jul 09 18:02:06 2004

Time range of VISIBLE PART:
  From 0.218114 to 1.210021 sec (duration 0.991907 sec)
Pitch:
  Median pitch: 122.813 Hz
  Mean pitch: 123.715 Hz
  Standard deviation: 5.165 Hz
  Minimum pitch: 115.129 Hz
  Maximum pitch: 136.464 Hz
Pulses:
  Number of pulses: 78
  Number of periods: 76
  Mean period: 8.099348E-03 seconds
  Standard deviation of period: 0.328449E-3 seconds
Voicing:
  Fraction of locally unvoiced frames: 26.174% (78 / 298)
  Number of voice breaks: 1
  Degree of voice breaks: 37.435% (0.371320 sec / 0.991907 sec)
Jitter:
  Jitter (local): 1.199%
  Jitter (local, absolute): 97.016E-6 seconds
  Jitter (rap): 0.546%
  Jitter (ppq5): 0.480%
Shimmer:
  Shimmer (local): 5.703%
  Shimmer (local, dB): 0.595 dB
  Shimmer (apq3): 2.550%
  Shimmer (apq5): 3.242%
  Shimmer (apq11): 5.337%
Harmonicity:
  Mean autocorrelation: 0.930093
  Mean noise-to-harmonics ratio (fraction): 0.098034
  Mean harmonics-to-noise ratio (dB): 14.873 dB

```

Fig. 4. Een stemrapport in Praat.

3. Stemhaperingen

De slechtste stemmen hebben moeite met het überhaupt laten dichtklappen van de stembanden. Figuur 3 toont een aangehouden klinker waarin een stempauze lijkt te vallen. In het stemrapport staan twee getallen die de mate van stemhapering aangeven. **Degree of voice breaks** geeft aan hoeveel procent van de tijd in beslag genomen wordt door stempauzes tussen de pulsen, d.w.z. intervallen tussen pulsen die langer zijn dan 1,25 maal de maximaal toegestane periode (in Fig. 3 is de minimaal toegestane toonhoogte 75 Hz, dus de maximaal toegestane periode $1/75 = 0,01333$ seconden, zodat intervallen van langer dan 0,01667 seconden als stemloos worden beschouwd). Zowel Praat als MDVP (met de maat **DVB**) geven hier 37% aan. Een andere maat is **Fraction of locally unvoiced frames**, die een onzichtbare eigenschap van de ‘ruwe’ toonhoogtemeting weergeeft. Deze tweede maat zegt waarschijnlijk meer over de stemkwaliteit dan de eerste, maar is sterk afhankelijk van de gebruikte toonhoogtemeetmethode. Zo geeft Praat hier een veel lager getal dan voor de eerste maat, terwijl MDVP (met de maat **DUV**) juist een veel hoger getal geeft. Ik raad u aan om **Degree of voice breaks** te gebruiken, vooral omdat die ook zichtbaar is in de pulsreeks van bv. Fig. 3.

Als **Degree of voice breaks** veel groter is dan **Fraction of locally unvoiced frames**, dan betekent dit dat Praat het stemloze stuk *een beetje* stemhebbend vond, maar niet stemhebbend genoeg om er stempulsen neer te zetten. Soms kan dit een

teken zijn dat er wel een goede toonhoogte aanwezig is, maar dat die onder de minimaal toegestane toonhoogte (in Fig. 3 is dat 75 Hz) ligt. In zo'n geval kunt u proberen om het stembereik opnieuw in te stellen (weer met **Pitch settings**). Praat's standaardinstelling is een bereik van 75 tot 500 Hz. Als voor het geluid van Fig. 3 de ondergrens wordt verlaagd tot 50 Hz, krijgen we Fig. 5. Inderdaad wordt nu het hele geluid als stemhebbend beschouwd. In het voorheen als stemloos beoordeelde stuk blijkt de spreker een toonhoogte van ongeveer 60 Hz te hebben, dus een octaaf lager dan in de rest van het geluid. Zulke *periodeverdubbelingen* komen veel voor in pathologische stemmen. Vaak (in Figuur 5 waarschijnlijk ook) is dit een teken van een onregelmatige regelmaat: de stembanden klappen wel degelijk 120 keer per seconde dicht, maar de even stembandklappen zien er anders uit dan de oneven. Dit kunt u controleren door in stukken met periodeverdubbelingen in te zoomen op een stuk of tien perioden van de golfvorm.

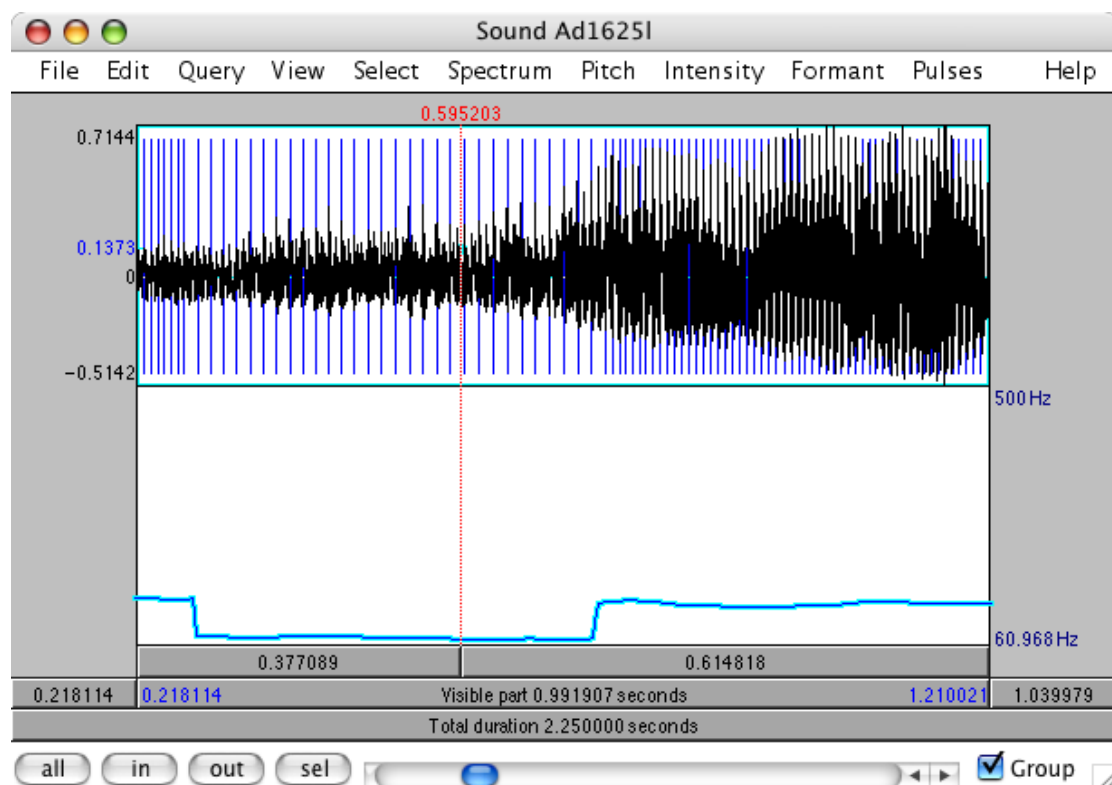


Fig. 5. Bij een andere instelling van het toonhoogtebereik blijkt er toch geen stempauze te zijn; wel een periodeverdubbeling.

U ziet dat u soms met het anders instellen van het stembereik de stemanalyse kunt verbeteren. Het geluid van Fig. 3 was speciaal uitgezocht om dit te kunnen illustreren, maar het komt regelmatig voor. Een toonhoogtecurve zou ook 'fouten' (uitschieters naar boven) kunnen maken. Die zijn vaak te verhelpen door de maximaal toegestane toonhoogte lager in te stellen dan 500 Hz. Voor mannenstemmen kunt u bijvoorbeeld dit maximum omlaag brengen tot 300 Hz. In alle gevallen 'krijgt u wat u ziet': de zichtbare pulsreeks is gebaseerd op de zichtbare toonhoogtecontour, en **Degree of voice breaks** weer op de zichtbare pulsreeks. Dit geldt ook voor de metingen die behandeld worden in de volgende paragrafen: jittermetingen worden gebaseerd op de zichtbare pulsreeks, shimmermetingen op de zichtbare pulsreeks en de zichtbare

golfvorm, en harmoniciteitsmetingen op de ‘kwaliteit’ van de zichtbare toonhoogtecontour.

Hoe goed is Praat in het meten van toonhoogte? Reetz (1996, p.163) vergeleek zes toonhoogteteetmethoden waarbij de methoden van Praat (Boersma 1993) en van Xwaves (Talkin 1995) als beste uit de bus kwamen. De methode van MDVP (Deliyski 1993) zat er overigens niet bij (die is volgens Dimitar Deliyski p.c. niet gelijk aan de methode van Kay/CSL).

4. Jitter

Als het de spreker niet lukt om opeenvolgende periodes een constante duur te geven, is er sprake van *jitter*. Figuur 6 toont hetzelfde geluid als in Figuur 1, maar voorzien van 10 procent ‘locale jitter’, hetgeen wil zeggen dat het gemiddelde verschil tussen twee opeenvolgende periodes 10 procent is.

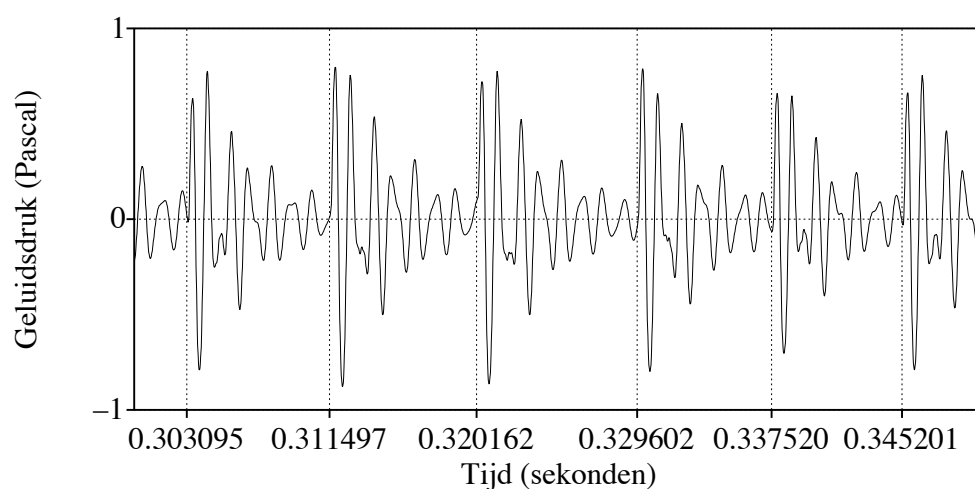


Fig. 6. Een geluid met 10 procent jitter.

Het geluid in Figuur 6 is een ‘synthetisch’ (met de computer gegenereerd) [a]-achtig geluid,¹ gebaseerd op een pulsreeks (een reeks tijdstippen) met bekende jitter. Figuur 6 toont de tijdstippen van de zes van deze onderliggende pulsen. De tweede periode duurt $0.320162 - 0.311497 = 0.008665$ seconden, de derde periode $0.329602 - 0.320162 = 0.009440$ seconden. Het verschil tussen deze twee opeenvolgende periodes is dus 0.000775 seconden. Aangezien het gemiddelde van de twee perioden 0.009057 seconden is, is de locale jitter $0.000775 / 0.009057 = 0.086$, ofwel 8,6 procent. In Figuur 6 (zes pulsen, vijf perioden) kunnen we op deze manier vier jitterwaarden uitrekenen. In Praat’s stemrapport staat het gemiddelde van deze jitterwaarden vermeld als **Jitter (local)**. Deze maat komt overeen met MDVP’s **Jitt**.

Een andere nuttige jittermaat is Praat’s **Jitter (rap)**, die overeenkomt met MDVP’s **RAP**. De afkorting staat voor *relative average perturbation*. Voor deze maat wordt elke periode vergeleken met de onmiddellijk voorafgaande periode en met de volgende periode. De eerste periode in Figuur 6 is $0,311497 - 0,303095 = 0,008402$ seconden.

¹ Technische details: om de [a]-kleuring te verkrijgen zijn de pulsen gefilterd met formanten van 820, 1300, 2300, 4200, 5400, 6600, 7800, 9000 en 10200 Hz. De eerste drie formanten hebben een bandbreedte die gelijk is aan de formantfrequentie gedeeld door 12, en de laatste zes een bandbreedte van 200 Hz.

Het gemiddelde van de eerste drie periodes is dus $(0,008402 + 0,008665 + 0,009440) / 3 = 0,008836$ seconden. De duur van de middelste (tweede) periode is iets lager dan dat gemiddelde, en wel $0,000171$ seconden, dat is $0,000171 / 0,008836 = 1,94$ procent van de gemiddelde periode. Deze RAP-waarden geven een beeld van de jitter als er sprake is van een bewegende toonhoogte: bij een jitterloze lineair stijgende toonhoogte is de *relative average perturbation* nul, terwijl de lokale jitter niet nul is. Voor een aangehouden klinker, waarbij de toonhoogte constant is, is de lokale jitter een goed maat. Zoals u kunt zien in het stemrapport kan deze ook in dB uitgedrukt worden. De vierde maat, Praat's **Jitter (ppq5)**, is alleen toegevoegd omdat MDVP de maat **PPQ** geeft (voor *pitch perturbation quotient*); de MDVP-handleiding geeft als reden dat deze maat minder gevoelig zou zijn voor fouten in de toonhoogtemeting. Voor Praat is dit irrelevant.

Hoe goed is de jittermeting? Om dit te kunnen bekijken werden geluiden met bekende jitter geanalyseerd met Praat en MDVP. De geluiden werden gegenereerd als boven beschreven en opgeslagen als 16-bit WAV-files met een bemonsteringsfrequentie van 44100 Hz (CD-kwaliteit, zoals vereist door MDVP). De geluiden werden met Praat geanalyseerd door de auteur van dit artikel, en met MDVP door Maria-Cristina Jackson-Menaldi (Wayne University, Detroit). Figuur 7a vergelijkt de prestaties van Praat en MDVP. Langs de horizontale as staat de werkelijke jitter van de onderliggende pulsreeksen, zoals rechtstreeks berekend uit de tijdstippen van de pulsen, langs de verticale as de door Praat en MDVP gemeten jitterwaarden.

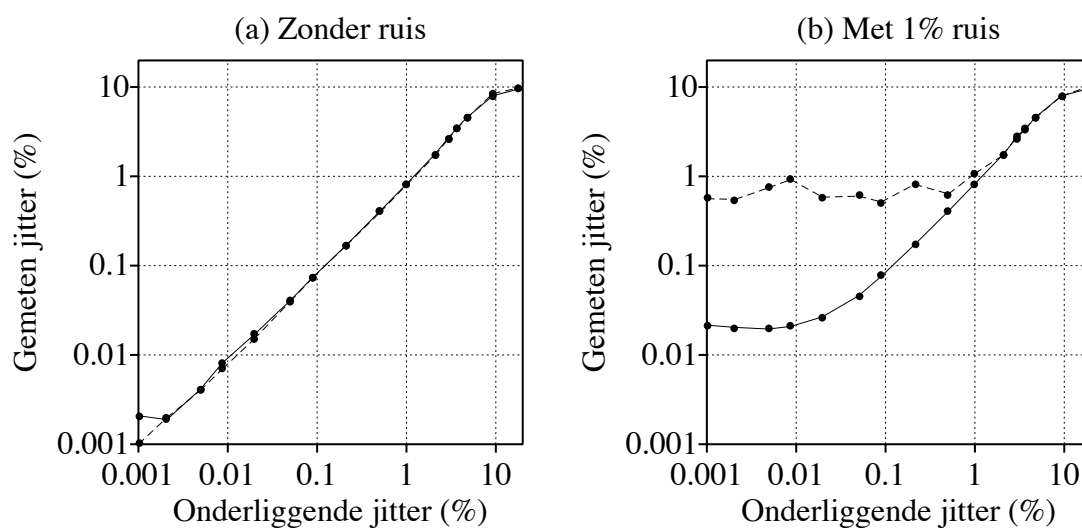


Fig. 7. Jittermetingen met Praat (—) en MDVP (- - -).

Aan het feit dat de curves diagonaal zijn zien we dat Praat en MDVP het beide heel goed doen: de gemeten jitter is ongeveer gelijk aan de onderliggende jitter. Wel liggen de curves iets onder de diagonaal die van de linkeronderhoek naar de rechterbovenhoek loopt, d.w.z. zowel Praat als MDVP onderschatten enigszins de onderliggende jitter. Dit komt door het nivellerende effect van doortrillende formanten. Boven de 10% laten zowel Praat als MDVP het afweten: een gemeten waarde van 10% kan zowel een onderliggende jitter van 10% als van 20% zijn. Dit komt doordat zowel Praat als MDVP zeer onregelmatige geluiden als niet-periodiek beschouwen, zodat voor een onderliggende jitter van 20% alleen die stukken gemeten worden die lokaal toevallig een relatief lage jitter hebben.

De klinische drempelwaarden voor jitter liggen rond de 1%. Deliyski (1993) geeft bijvoorbeeld 1,04% als de grenswaarde tussen 15 normale sprekers en 53 patiënten met vijf verschillende pathologieën, gemeten voor een volgehouden [a]. Figuur 7a laat zien dat zowel Praat als MDVP zeer goed presteren in het klinisch relevante gebied.

De prestaties van de programma's veranderen dramatisch als we ruis toevoegen. In Figuur 7b is witte ruis toegevoegd met een vermogen van 1% van het vermogen van het 'signaal' (de gefilterde pulstrein); zo'n geluid is te zien in Figuur 8. Een ruisvermogen van 1% is niet zo groot: veel patiënten hebben een stuk hogere additieve ruis dan 1% (d.w.z. een stuk lagere *HNR* dan 20 dB, zie §6). Toch gaat bij deze niet zo abnormale additieve ruis van 1% het programma MDVP de mist in: de gemeten jitter voor een jittervrij geluid zal rond de 0,5% liggen, wat vergelijkbaar is met de door Hillenbrand (1987) gemelde meetwaarde voor jittervrije geluiden (0,6%). Ook zien we in Figuur 7b dat het mogelijk is dat MDVP de jitter van een geluid met een onderliggende jitter van 0,01% als 1% kan meten. Meetwaarden rond de 1%, dus in het klinisch relevante gebied, zijn dus in MDVP onbetrouwbaar als er sprake is van zelfs maar een matige ruis van 1%. Op deze gronden zegt Hillenbrand (1987) dat jittermetingen slechts als 'generieke metingen' van aperiodiciteit moeten worden beschouwd, met weinig verband met werkelijke onregelmatigheden in de duur van stemperiodes. Maar zo pessimistisch hoeven we tegenwoordig niet meer te zijn. Praat meet namelijk bij een ruis van 1% betrouwbare jitterwaarden in het hele gebied tussen de 0,1% en de 10% jitter, zoals Figuur 7b laat zien. Voor hogere ruiswaarden dan 1% worden de jittermetingen in beide programma's onbetrouwbaarder, ofschoon Praat zelfs bij een ruis van 100% (d.w.z. even veel ruis als 'signaal') een jitter van slechts 0,4% meet. Wil een klinische beslissing op basis van jittermetingen dus betrouwbaar zijn, dan moet de additieve ruis beneden de 100% zijn in Praat, of beneden de 1% in MDVP.

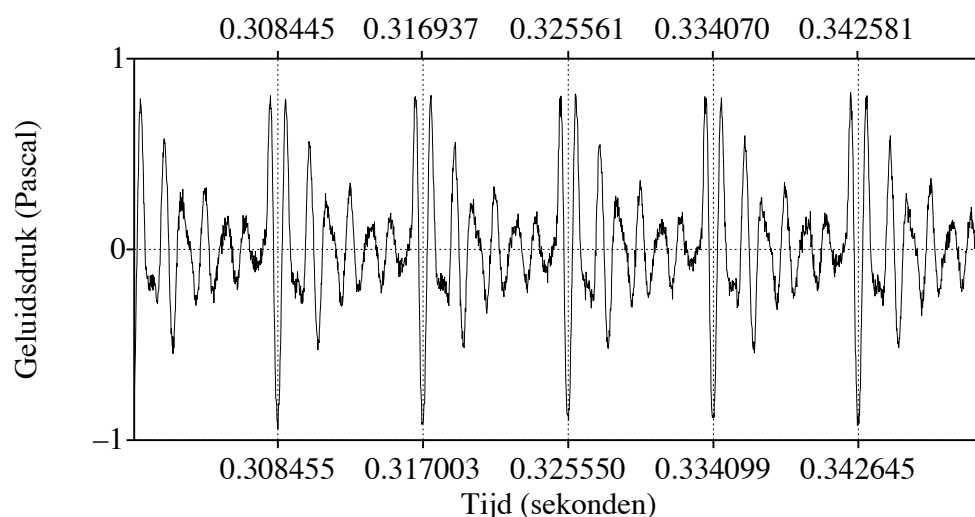


Fig. 8. Het geluid van Figuur 1 met 1% additieve witte ruis.
Langs de onderrand: pulsmetingen door gelijkheid van golfvormen;
langs de bovenrand: pulsmetingen door hoogte van pieken.

Het verschil tussen Praat en MDVP wordt veroorzaakt door de methode van bepaling van de tijdstippen van de pulsen. De onderrand van Figuur 8 toont de zes tijdstippen die door Praat bepaald zijn op basis van patroonherkenning: de golfvorm rond 0,317003 seconden lijkt meer op de golfvorm rond 0,325550 seconden dan bv. op de

golfvorm rond 0,325000 of 0,326000 seconden. De vier aldus gemeten periodes, af te lezen uit de getallen onder Figuur 8, zijn 8,548 ms, 8,547 ms, 8,549 ms en 8,546 ms, allemaal zeer dicht bij elkaar: de gemeten jitter over het hele geluid is slechts 0,02%. MDVP daarentegen meet de tijdstippen van de hoogste (of diepste) pieken in het geluid. Bij de aanwezigheid van ruis is de kans groot dat het tijdstip van een piek gedeeltelijk door een toevallige ruispiek wordt bepaald. Zulke tijdstippen zijn langs de bovenrand van Figuur 8 aangegeven (zoals U zelf in Praat zou kunnen meten door een geluid te selecteren en dan **To PointProcess (periodic, peaks)** te kiezen). De periodes die zijn af te lezen uit deze getallen zijn 8,492 ms, 8,624 ms, 8,509 ms en 8,511 ms, duidelijk een stuk verder uit elkaar. De gemeten jitter over het hele geluid is hier 0,6%, dezelfde waarde als MDVP opgeeft voor dit geluid. Dit verklaart de grote gevoeligheid van de MDVP-jittermeting voor additieve ruis.

5. Shimmer

Als het de spreker niet lukt om opeenvolgende periodes een constante amplitude te geven, is er sprake van *shimmer*. Figuur 9 toont hetzelfde geluid als in Figuur 1, maar voorzien van 10 procent 'locale shimmer', hetgeen wil zeggen dat het gemiddelde verschil tussen de amplitudes van twee opeenvolgende periodes 10 procent is.

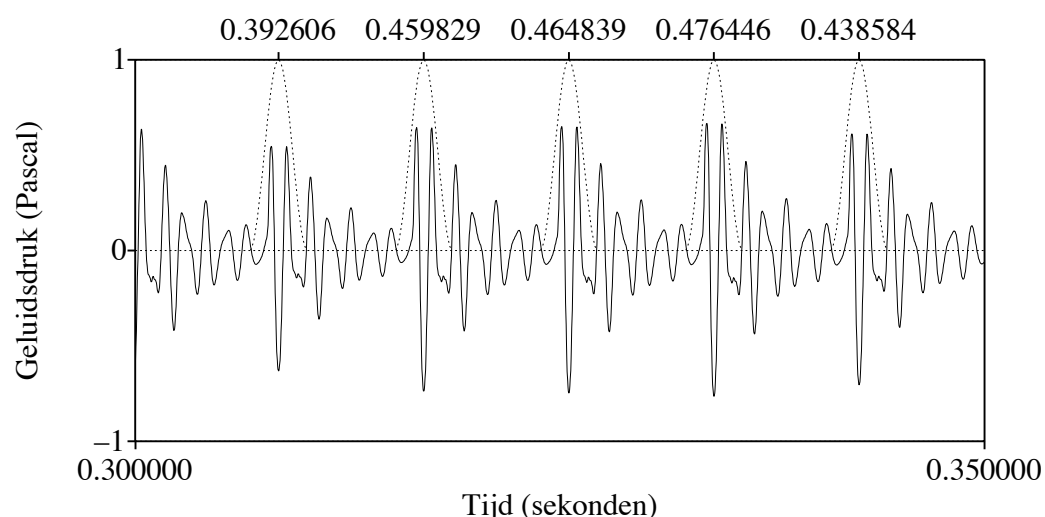


Fig. 9. Een geluid met 10 procent shimmer.

Praat meet de shimmer op basis van de energie rond de tijdstippen in het signaal die Praat's pulsmeet algoritme gevonden heeft (zoals langs de onderrand van Figuur 8). De getallen langs de bovenrand in Figuur 9 geven de 'rms' (root-mean-square) amplitude in de bijbehorende klokvormige stukjes signaal (de wortel van de gemiddelde energie). Het verschil tussen 0,392606 en 0,459829 is groot. De bijdrage van dit verschil aan de locale shimmer wordt als volgt berekend: de gemiddelde amplitude is $(0,392606 + 0,459829) / 2 = 0,4262175$ Pa; het verschil tussen de amplitudes is $0,459829 - 0,392606 = 0,067223$ Pa; relatief is dat $0,067223 / 0,4262175 = 15,8\%$. Evenzo is het verschil tussen 0,459829 en 0,464839 ongeveer 1,1%. Gemiddeld over het hele geluid meet Praat een locale shimmer van 8,6%. Over het algemeen komen de shimmermetingen van Praat (gemeten door de auteur dezes) en MDVP (gemeten door Maria Cristina Jackson-Menaldi) sterk overeen, zoals Figuur 10a laat zien: in het gebied tussen 0.1% en 10% zijn de metingen betrouwbaar.

MDVP maakt voor de shimmermetingen gebruik van de pieken in het geluid, d.w.z. in Figuur 9 meet MDVP alleen de hoogte van de negatieve pieken rond de tijdstippen van de pulsen. Net als bij de jittermetingen kunnen we dan verwachten dat de shimmermetingen in MDVP gevoeliger zijn voor additieve ruis dan die van Praat. Figuur 10b laat zien dat dit inderdaad zo is. Met slechts 1% additieve ruis meet MDVP al een shimmer van 2%, ook als de werkelijke onderliggende shimmer nul is; dit getal van 2% is niet zoveel beter dan de door Hillenbrand (1987) gemelde waarde van 0,26 dB (= 3%). Deze hoge meetfout is lastig, want de drempelwaarde (tussen gezond en pathologisch) van 3,81% (Deliyski 1993) ligt er niet ver boven. Net als bij de jittermetingen concludeert Hillenbrand (1987) dan ook dat shimmermetingen alleen generieke periodiciteitsafwijkingen meten en niet direct gerelateerd zijn aan werkelijke onregelmatigheden in de amplitudes van stemperiodes. Alweer is zulk pessimisme tegenwoordig ongepast. Zoals Figuur 10b laat zien, meet Praat hier immers een shimmer van slechts 0,7%. Voor hogere waarden van additieve ruis zijn de gemeten shimmerwaarden voor beide programma's hoger: bij een ruis van 100% meet MDVP een shimmer van 10%. Praat meet 2,4% bij een ruis van 10%, 5,8% bij een ruis van 100%. Wil een klinische beslissing op basis van shimmermetingen dus betrouwbaar zijn, dan moet de additieve ruis beneden de 10% zijn in Praat, of zelfs beneden de 1% in MDVP.

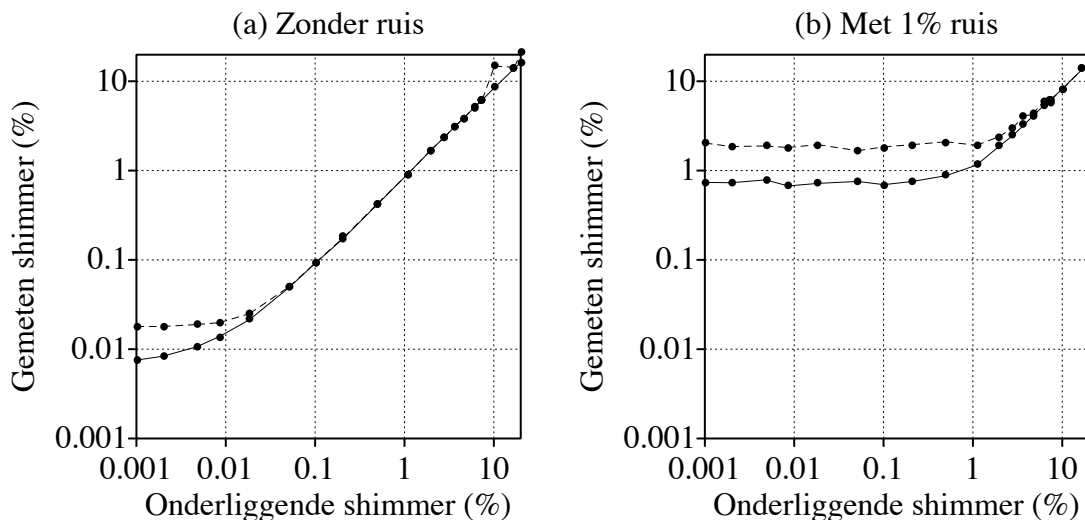


Fig. 10. Shimmermetingen met Praat (—) en MDVP (- - -).

6. 'Ruizigheid' (harmonicity)

Additieve ruis beïnvloedt niet alleen de jitter- en shimmermetingen, maar kan ook zelf gemeten worden, althans door Praat. In Figuur 4 zien we onder het kopje **Harmonicity** de **Mean autocorrelation**, een maat voor de mate waarin de golfvormen van twee opeenvolgende periodes op elkaar lijken. Bij een perfect periodiek signaal is deze gemiddelde autocorrelatie 1, voor witte ruis ongeveer 0. De autocorrelatie wordt beïnvloed door jitter, shimmer en additieve ruis. Voor een signaal zonder jitter en shimmer maar met 1% additieve ruis, zoals het geluid in Figuur 8, zal Praat een gemiddelde autocorrelatie van ongeveer 99% rapporteren (de verwachte waarde is 100/101). Een tweede maat is de **Mean noise-to-harmonics ratio (fraction)**. Voor elke periode wordt de ruis-harmonischenverhouding

rechtstreeks uit de autocorrelatie bepaald: het is 1 minus de autocorrelatie. Voor het geluid van Figuur 8 is deze verhouding 0.01 (= 1% ruis gedeeld door 100% signaal). Een vergelijkbare maat (de ruisenergie tussen 1500 en 4500 Hz gedeeld door de harmonische energie tussen 70 en 4500 Hz) komt ook voor in het stemrapport van MDVP en heet daar **NHR**, maar helaas kan dat programma nauwelijks lagere waarden dan 0.1 meten. Een derde maat is de **Mean harmonics-to-noise ratio (dB)**. Voor elke periode kan deze *HNR* berekend worden als -10 maal de logaritme van de ruis-harmonischenverhouding. Voor het signaal van Figuur 8 (1% ruis) is dat 20 dB.

Hoe goed is Praat in het meten van HNR? Figuur 11 laat de gemeten HNR zien als functie van de werkelijke HNR. De geluiden die gemeten werden voor Figuur 11 zijn dezelfde synthetische geluiden als in Figuur 8, maar met een variërende additieve ruis. We zien dat Praat geen moeite heeft met HNR-waarden tussen 0 dB en 65 dB (d.w.z. ruis-harmonischenverhoudingen tussen 0,0000003 en 1), terwijl MDVP slechts woorden tussen 0 dB en 10 dB kan meten (d.w.z. ruis-harmonischenverhoudingen tussen 0,1 en 1). Praat meet ruis dus 300.000 keer zo precies als MDVP; de oorzaak hiervan is Praat's patroonherkenningsmethode ("CC", voor *cross-correlation*, die U krijgt met de instelling **Optimize for voice analysis**, zie §2), die een combinatie is van de kruiscorrelatiemethode van Talkin (1995) en de sinc-interpolatiemethode van Boersma (1993). Als U de patroonherkenningsmethode op "AC" (= *autocorrelation*) zet (door de instelling **Optimize for intonation** te kiezen, zie §2) gaat de maximale betrouwbare HNR-waarde omlaag naar 30 dB, wat overigens nog goed genoeg is voor spraakpathologisch onderzoek (ofschoon niet goed genoeg voor de [u] van auteur dezes, die een HNR van 40 dB heeft).

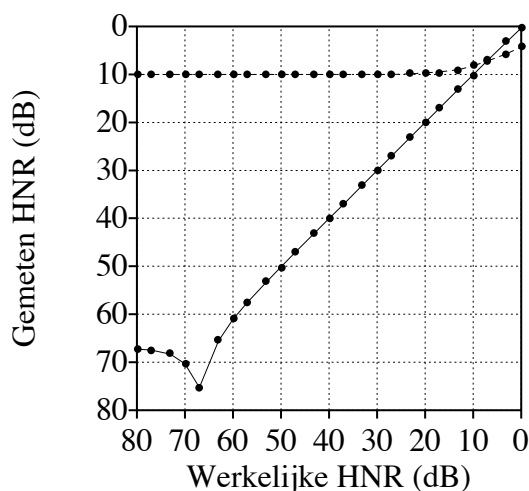


Fig. 11. Ruismetingen met Praat (—) en MDVP (- - -).

Overigens moet gezegd worden dat Hillenbrand's (1987) pessimisme over HNR-metingen, anders dan zijn pessimisme over jitter- en shimmermetingen, wèl gerechtvaardigd is: gemeten HNR-waarden weerspiegelen niet alleen de additieve ruis, maar ook jitter en shimmer, en zijn dus geen goed diagnostisch criterium voor een specifieke afwijking van de periodiciteit.

7. Conclusie

Het aantal verschillende stemmeetmethoden dat U redelijkerwijs van een spraakanalyseprogramma, en dus ook van Praat, kunt verwachten, is vreemd genoeg

afhankelijk van de kwaliteit van de stem. Voor heel slechte stemmen zijn alleen stempauzemetingen mogelijk (omdat toonhoogteteetmethoden daar veel “fouten” maken). Voor iets betere stemmen zijn met Praat metingen van ruis (HNR), jitter en shimmer mogelijk, hoewel de ruis voor shimmermetingen onder de 10% moet zijn. Voor MDVP geldt dat ruismetingen nauwelijks mogelijk zijn, en jitter- en shimmermetingen alleen als de ruis minder is dan 1% (hetgeen in MDVP dus niet is vast te stellen); in een rechtstreekse vergelijking met spraakachtige synthetische signalen blijken de jittermetingen van Praat minstens 10 keer minder gevoelig voor ruis dan die van MDVP, en blijkt Praat’s HNR-meetbereik 55 dB groter dan dat van MDVP.

Een stemrapport maakt u in Praat na het selecteren van een relevant gedeelte van het geluid. Als het erom gaat te beoordelen of de patiënt in staat is een klinker aan te houden, dan is het raadzaam om het begin en het eind van het geluid niet mee te selecteren, omdat hier zelfs bij gezonde stemmen onregelmatigheden in kunnen zitten die de metingen ongunstig beïnvloeden.

Het programma Praat is vrij te downloaden van www.praat.org.

Dank

Mijn dank gaat vooral uit naar Maria Cristina Jackson-Menaldi, die mijn 79 synthetische geluiden analyseerde met MDVP en het geluid van Figuur 2 tot 5 beschikbaar stelde. Verder dank aan Petra Jongmans, Frans Hilgers en Corina van As voor opmerkingen over een eerdere versie van dit artikel.

Referenties

- As, Corina van (2001). *Tracheoesophageal speech: a multidimensional assessment of voice quality*. Doctoral thesis, University of Amsterdam.
- Boersma, Paul (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetics Sciences of the University of Amsterdam*, 17, 97–110.
- Deliyski, Dimitar D. (1993). Acoustic model and evaluation of pathological voice production. *Proceedings Eurospeech '93*, Berlin. 1969–1972.
- Hillenbrand, James (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. *Journal of Speech and Hearing Research* 30, 448–461.
- Iwata, Shigenobu, en Leden, H. von (1970). Pitch perturbations in normal and pathologic voices. *Folia Phoniatica* 22, 413–424.
- Klingholz, F., en Martin, F. (1983). Speech wave aperiodicities at sustained phonation in functional dysphonia. *Folia Phoniatica* 35, 322–327.
- Koike, Yasuo (1969). Vowel amplitude modulations in patients with laryngeal diseases. *Journal of the Acoustical Society of America* 45, 839–844.
- Lieberman, Philip (1963). Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *Journal of the Acoustical Society of America* 35, 344–353.
- Reetz, Henning (1996). *Pitch perception in speech: a time domain approach*. Doctoral thesis, University of Amsterdam. Amsterdam: IFOTT.
- Talkin, David T. (1995). A robust algorithm for pitch tracking. In W. Bastiaan Kleijn & Kuldip K. Paliwal (eds.), *Speech coding and synthesis*. Amsterdam: Elsevier. 495–518.
- Verdonck-De Leeuw, Irma (1998). *Voice characteristics following radiotherapy: the development of a protocol*. Doctoral thesis, University of Amsterdam.
- Wendahl, R.W. (1966). Laryngeal analog synthesis of jitter and shimmer, auditory parameters of hoarseness. *Folia Phoniatica* 18, 98–106.