

Project BIG

A. (Arthur) van Dalen

BSc Report

Committee:

Prof.dr.ir. C.H. Slump
Dr.ir. J.B.C. Engelen
Dr.ir. F. van der Heijden

November 2018

042RAM2018
Robotics and Mechatronics
EE-Math-CS
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Voorwoord

Dit verslag is geschreven als eindproduct voor de bachelor drie opdracht van de opleiding Electrical Engineering aan de Universiteit Twente. Het is een haalbaarheidsonderzoek voor een methode om biggensterfte te reduceren in de varkenshouderij.

Enschede, november 2018

Arthur van Dalen

Samenvatting

In dit onderzoek wordt de mogelijkheid onderzocht om met behulp van audiosignalen een noodkreet van een jonge big te detecteren. Hiervoor zijn audiofragmenten uit een biggenstal met zeugen en hun pasgeboren biggen beschikbaar gesteld. Uit deze audiofragmenten worden de noodkreten geïsoleerd en geanalyseerd. Eerst worden sterk stationaire ruisbronnen, die buiten het spectrum van de noodkreet vallen, geëlimineerd met behulp van filtering. Dan wordt er onderzocht of er met de korte termijn eigenschappen van het audiofragment genoeg onderscheid gemaakt kan worden tussen een noodkreet en de andere geluiden in de stal. Bij audio identificaties worden veelal de volgende vijf eigenschappen gebruikt: de short term energy, de zero crossing rate, de spectral centroid, de roll of point en de Mel-frequency cepstrum coefficients. Deze eigenschappen zullen onderzocht worden in de gevallen dat er een noodkreet is, als het vrij stil is in de stal en als de biggen veel lawaai maken terwijl er geen nood is. Vervolgens wordt er onderzocht welk herkenningsmodel geschikt is om te gebruiken in deze situatie. Er is gekozen om de gaussian mixture model, de support vector machine en de K-nearest neighbor methode te testen. Deze methodes dienen getraind te worden met behulp van de korte termijn eigenschappen van de audiofragmenten. Om te onderzoeken hoeveel de kwaliteit van het opname materiaal invloed heeft op de classificatie, wordt er ook een model gemaakt van de korte termijn eigenschappen van een noodkreet. Uiteindelijk zal onderzocht moeten worden of de methodes rendabel kunnen zijn in een varkenshouderij. Hiervoor zijn in overleg met de opdrachtgever een aantal eisen opgesteld waaraan de methodes moeten voldoen, deze eisen zijn gebaseerd op precisie in het detecteren van noodsituaties en de benodigde rekenkracht.

Inhoudsopgave

1	Inleiding	9
2	Vooronderzoek	11
2.1	Beschikbaar onderzoeksmateriaal	11
2.2	Audiokarakteristieken	11
2.2.1	Short term energy	11
2.2.2	Zero crossing rate	12
2.2.3	Spectral centroid	12
2.2.4	Roll off point	12
2.2.5	Mel-frequency cepstrum coefficients (MFCC)	13
2.3	Herkennings Modellen	14
2.3.1	Gaussian Mixture Model	14
2.3.2	Support vector machine	14
2.3.3	K-nearest neighbor	15
2.4	Voorbewerkingen	15
3	Onderzoeksmethode	17
3.1	Probleemstelling	17
3.2	Doelstellingen	17
3.3	Centrale onderzoeksvraag	17
3.3.1	Subvragen	17
4	Signaal analyse	19
4.1	Analyse Doodskreet	19
4.1.1	Short term energy	20
4.1.2	Zero crossing rate	21
4.1.3	Spectral centroid	22
4.1.4	Roll off point	23
4.1.5	Mel-frequency cepstrum coefficients (MFCC)	23
4.2	Het systeem	24
5	Model Training	25
5.1	Feature selectie	25
5.1.1	Feature modellering	25
5.1.2	Testsignaal	28
5.2	Gaussian mixture model	29
5.2.1	Model training	29
5.2.2	Model training met behulp van gemodelleerd signaal	29
5.3	Support vector machine	31
5.3.1	Model training	31
5.3.2	Model training met behulp van gemodelleerd signaal	31
5.4	K-nearest neighbour	34
5.4.1	Model training	34
5.4.2	Model training met gemodelleerd signaal	34
6	Conclusie	39
6.1	Resultaten tabel	39
6.2	Receiver operating characteristic figuur	40
6.3	Conclusie	41
6.4	Aanbevelingen	41
7	Literatuurlijst	43

8	Bijlage	45
8.1	Bijlage 1: Karakteristieken van de Mel-frequency cepstrum coefficients (MFCC)	47
8.2	Bijlage 2: Uiteenzetting van geluidsfeatures	51

Lijst van figuren

1	Frequentie responsie van een Mel filterbank	13
2	Tijd/frequentie plot van een doodskreet, het lichter gebied bevat meer energie en de hoogte geeft de frequentie aan.	19
3	Uitvergroting van de tijd/frequentie plot, geconcentreerd op het gebied dat een doodskreet bevat	19
4	Tijd/frequentie plot van een audiofragment dat alleen achtergrond ruis bevat.	20
5	Een audiofragment van een doodskreet, in blauw aangegeven, met de bijbehorende korte termijns energie aangegeven in rood	21
6	Een audiofragment van de achtergrond ruis, in blauw aangegeven, met de bijbehorende korte termijns energie aangegeven in rood	21
7	Een audiofragment van een doodskreet, aangegeven in blauw, met de bijbehorende zero crossing rate aangegeven in rood	22
8	Een audiofragment van de achtergrond ruis, aangegeven in blauw, met de bijbehorende zero crossing rate aangegeven in rood	22
9	Een audiofragment van een doodskreet, aangegeven in blauw, met de bijbehorende spectral centroid aangegeven in rood	22
10	Een audiofragment van achtergrond ruis, aangegeven in blauw, met de bijbehorende spectral centroid aangegeven in rood	23
11	Een audiofragment van een doodskreet, aangegeven in blauw, en de bijbehorende roll off point aangegeven in rood	23
12	Een audiofragment van achtergrondruis, aangegeven in blauw, en de bijbehorende roll off point aangegeven in rood	23
13	Systematische weergave van het alarmsysteem	24
14	Short term energy van een doodskreet, weergegeven in groen, en van achtergrondruis, weergegeven in rood, uitgezet tegen de zero crossing rate van dezelfde signalen	26
15	Short term energy van het model signaal uitgezet tegen de Zero crossing rate van het model signaal	27
16	Figuur 14 en figuur 15 samengevoegd	27
17	Testsignaal opgebouwd met audiofragmenten van achtergrondruis, een doodskreet en ruziënde biggen	28
18	Testsignaal, aangegeven in blauw, met het resultaat van een gaussian mixture model meting, aangegeven in rood, met de short term energy en de spectral centroid als features	29
19	Testsignaal, aangegeven in blauw, met het resultaat van een gaussian mixture model meting, aangegeven in rood, met de short term energy en de spectral centroid van het modelsignaal	30
20	Support vector machine resultaten, aangegeven door omcirkeling, met de spectral centroid en de shortterm energy van achtergrondruis, aangegeven in rood, en een doodskreet, aangegeven in groen, tegen elkaar uitgezet	32
21	Testsignaal, aangegeven in blauw, met het resultaat van de support vector machine, aangegeven in rood, met de zero crossing rate en de short term energy als features	32
22	Testsignaal, aangegeven in blauw, met het resultaat van support vector machine, aangegeven in rood, met zero crossing rate en short term energy van het modelsignaal als features	33
23	Testsignaal, aangegeven in blauw, met resultaat van K-nearest neighbour, aangegeven in rood, met spectral centroid, short term energy, zero crossing rate en 1 vergelijkingspunt	35

24	Testsignaal, aangegeven in blauw, met resultaat van K-nearest neighbour, aangegeven in rood, met spectral centroid, short term energy, zero crossing rate en 10 vergelijkingspunten	36
25	Testsignaal, aangegeven in blauw, met resultaat van K-nearest neighbour, aangegeven in rood, met spectral centroid, short term energy, zero crossing rate en 100 vergelijkingspunten	37
26	Testsignaal, aangegeven in blauw, met resultaat van K-nearest neighbour, aangegeven in rood, met spectral centroid, short term energy, zero crossing rate van het modelsignaal en 100 vergelijkingspunten	38
27	Receiver operating characteristic van de verschillende methoden	40
28	Uitvergroting meetpunten receiver operating characteristic figuur	40
29	Audio opname van een doodskreet, weergegeven in blauw, met de eerste 4 MFCC karakteristieken, weergegeven in rood	47
30	Audio opname van een doodskreet, weergegeven in blauw, met MFCC karakteristieken 5 tot en met 8, weergegeven in rood	48
31	Audio opname van een doodskreet, weergegeven in blauw, met MFCC karakteristieken 9 tot en met 13, weergegeven in rood	48
32	Audio opname van achtergrondgeluid van een stal, weergegeven in blauw, met MFCC karakteristieken 1 tot en met 4, weergegeven in rood	49
33	Audio opname van achtergrondgeluid van een stal, weergegeven in blauw, met MFCC karakteristieken 5 tot en met 8, weergegeven in rood	49
34	Audio opname van achtergrondgeluid van een stal, weergegeven in blauw, met MFCC karakteristieken 9 tot en met 13, weergegeven in rood	50
35	De spectral centroid uitgezet tegen de zero crossing rate	51
36	De spectral centroid uitgezet tegen de MFCC 1 setting	51
37	De shortterm energy uitgezet tegen de spectral centroid	52
38	De shortterm energy uitgezet tegen de MFCC coefficienten	52
39	de zero crossing rate uitgezet tegen de MFCC coefficienten	52

1 Inleiding

Wanneer een zeug biggetjes heeft geworpen, worden deze biggetjes ook gezoogd door de zeug. In deze situatie dient de zeug te gaan liggen en hierbij komt het wel eens voor dat de zeug op een biggetje is gaan liggen. Pas geboren biggetjes zijn niet sterk genoeg om zelf te ontsnappen en kunnen dood gedrukt worden door de zeug. Het is een natuurlijke reactie van een big om heel hard te gaan schreeuwen bij beklemming, waardoor de zeug weer op zal gaan staan. In een significant aantal gevallen staat de zeug niet op en blijft het big klem zitten, dit kan leiden tot zwaar letsel of sterfte bij de big. In de varkenshouderij is elke dood gedrukte big een verlies van inkomsten. Het is mogelijk om een systeem te ontwikkelen wat alarmeert zodra een big klem zit, zodat de big gered kan worden. In dit onderzoek wordt een detectie methode onderzocht voor het detecteren van een big die vastzit onder een zeug. Het medium dat hier gebruikt wordt is geluid. Er wordt onderzocht of de schreeuwen van een big in nood kunnen worden onderscheiden van de rest van de geluiden in een stal. Hiervoor dienen eerst de eigenschappen van een doodskreet geanalyseerd te worden. Aan de hand van deze eigenschappen zal dan een detectiemethode ontworpen worden. De detectiemethode wordt dan beoordeeld op het aantal doodskreten die succesvol worden gedetecteerd en het aantal valse meldingen en of dit aantal aan de gestelde eisen voldoet.

2 Vooronderzoek

Eerst wordt voor dit onderzoek informatie vergaard over het detecteren van specifieke geluiden. Een soortgelijke onderzoek over het lokaliseren en identificeren van het hoesten van biggen, het werk van Exadaktylo, Silva, Ferrari, Guarino, Taylor, Aerts en Berkmans (2009)⁴, kan als uitgangspunt dienen voor dit onderzoek. Als vergelijk kan het werk van Drugman (2014)¹⁵ over het detecteren en classificeren van het hoesten van mensen gebruikt worden, aangezien dit om hetzelfde soort geluid gaat. Informatie uit het onderzoek van Xie, Towsey, Zhang en Roe (2016)⁶ over het identificeren van verschillende soorten kikkers door middel van hun roepen kan gebruikt worden om onderscheidt te maken van de verschillende biggen kreten. De belangrijke informatie uit deze bronnen is de methode waarmee een specifieke kreet van een dier gedetecteerd kan worden, zodat er een methode voor het detecteren van een doodskreet van een big ontworpen gerealiseerd kan worden. Bij deze methodes wordt eerst het audiosignaal uiteengezet in een aantal karakteristieke eigenschappen, deze zullen dan worden onderzocht om de geschiktheid te bepalen voor het identificeren van een doodskreet. Voor het identificeren en daarna classificeren van audiosignalen worden voornamelijk lerende machine methodes gebruikt. Ook wordt er onderzocht welke voorbewerkingen kunnen worden toegepast op het audiosignaal voor het optimaliseren van het berekenen van de audiofeatures.

2.1 Beschikbaar onderzoeksmateriaal

Voor dit onderzoek zijn er gedurende drie dagen geluidsopnamen in een stal met pasgeworpen biggetjes gemaakt. Er is opgenomen met acht microfoons die verdeeld zijn over de stal. Deze opname is gesegmenteerd in audiobestanden die een half uur duren. Er is ook een lijst beschikbaar van de tijdstippen wanneer een big klem zit onder een zeug.

2.2 Audiokarakteristieken

Bij de meeste vormen van stemherkenning wordt er gebruik gemaakt van de korte termijn eigenschappen van het audiosignaal, zie het werk van Babaei, Anuar, Wahab en Shamshirband (2017)³ en het werk van Li (2000)¹². Er is gekozen om de volgende korte termijn eigenschappen te onderzoeken:

- Short term energy
- Zero crossing rate
- Spectral centroid
- Roll off point
- Mel-frequency cepstrum coefficients

Deze vijf eigenschappen worden veel gebruikt bij audio identificatie methodes. Er wordt onderzocht wat deze vijf eigenschappen zijn en later zal onderzocht worden of deze eigenschappen toepasbaar zijn op de probleemstelling van dit project.

2.2.1 Short term energy

De korte termijn energie is de hoeveelheid energie die in het audiosignaal zit. De hoeveelheid energie varieert aan de hand van het signaal. Ruis heeft een constante

hoeveelheid vermogen en bij het geluid van iemand die praat fluctueert het vermogen met de tijd. De berekening voor deze eigenschap is gegeven in vergelijking 1. In deze vergelijking is x het audiosignaal en w de toegepaste Hamming window voor het korte termijn karakteristiek. De Hamming window is gekozen, omdat deze de signalen buiten de band beter onderdrukt.

$$E_m = \sum_{n=0}^{\infty} [x(n) w(m-n)]^2 \quad (1)$$

2.2.2 Zero crossing rate

Deze eigenschap is het aantal keer dat het signaal van teken wisselt. Bij een audio signaal zonder bias is dit wanneer de amplitude van het signaal 0 passeert. Bij een discreet signaal is dit het aantal keer dat het teken omklapt. Met deze eigenschap kan de frequentie van het signaal geschat worden. Bij stemherkenning varieert de frequentie van het signaal over de tijd en kan deze eigenschap gebruikt worden om het te detecteren. De zero crossing rate van een discreet signaal kan berekend worden met de formule in vergelijking 2. In deze vergelijking is de functie $w(m-n)$ toegepast als rechthoekig raamwerk.

$$Z_m = \sum_{n=0}^{\infty} |\text{sign}[x(n)] - \text{sign}[x(n-1)]| w(m-n) \quad (2)$$

$$\text{sign}[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases} \quad (3)$$

2.2.3 Spectral centroid

De spectral centroid is het gewogen gemiddelde van het frequentiespectrum. Bij achtergrond ruis zal deze eigenschap constant blijven, omdat de meest aangetroffen ruis soorten uniform het vermogen hebben verdeeld over het volledig spectrum. Als er een stem of in het geval van dit project een big start met schreeuwen, zal het zwaartepunt verschuiven naar het frequentie gebied van het geluid. De spectral centroid wordt berekend met vergelijking 4 hierbij is f de centrale frequentie en g is de wegingsfactor. De wegingsfactor wordt berekend door de fourier transformatie van het audiosignaal te nemen en dan de hoogte van de passende frequentie te nemen, voor het bepalen van de Fourier transformatie zie Haykin en Moher hoofdstuk 2¹⁰.

$$\text{Centroid} = \frac{\sum f(n) g(n)}{\sum g(n)} \quad (4)$$

2.2.4 Roll off point

De roll off point is de frequentie in een signaal wanneer de hogere frequenties nog maar 15 % van het totale energie in het signaal bevatten. In het geval dat het signaal voornamelijk uit ruis bestaat, zal de roll of point op 85 % van het totale spectrum zitten. Maar als er een ander geluid in het signaal bijkomt, zoals een stem of een schreeuw van een big, dan zal het roll of point veranderen. Dit komt, omdat er een hogere concentratie van de totale signaal energy zich op een andere frequentie bevindt dan voorheen. Het roll off point wordt berekend met vergelijking 5 hierbij is R de roll of frequentie, N_m is de maximale frequentie van het signaal en $f(n)$ is de power spectral density van het signaal, voor de bepaling van de power spectral density zie Haykin en Moher hoofdstuk 2¹⁰.

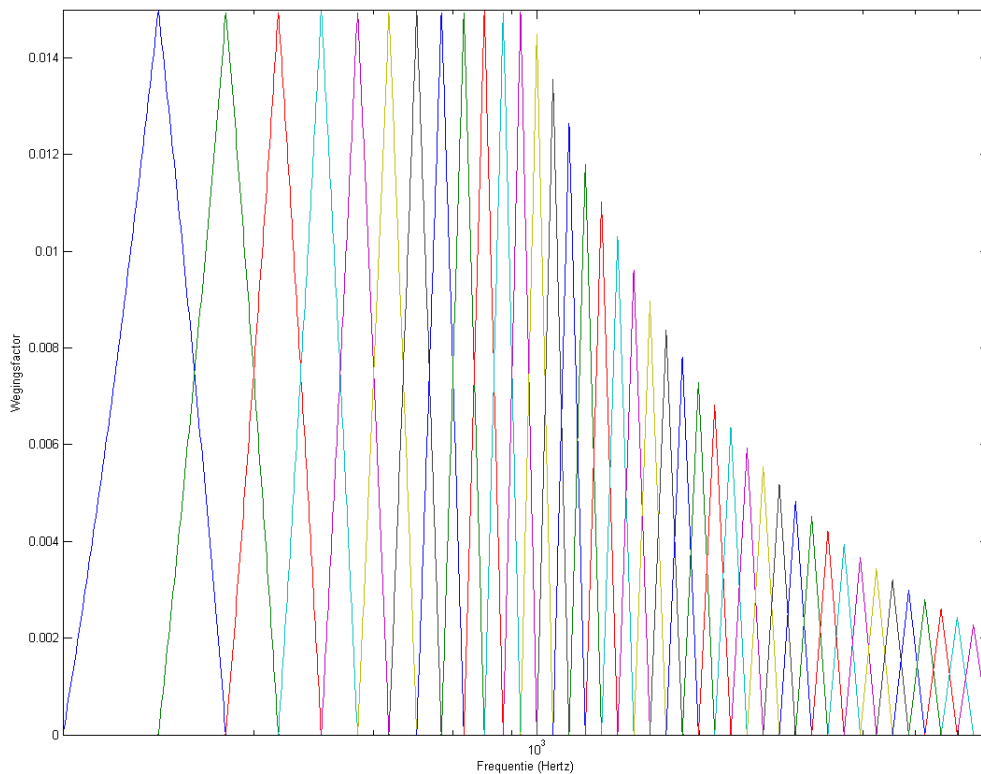
$$\text{Rolloff point} \rightarrow \sum_{n=1}^R f(n) = 0.85 \sum_{n=1}^{N_m} f(n) \quad (5)$$

2.2.5 Mel-frequency cepstrum coefficients (MFCC)

Een cepstrum is de Fourier transformatie van het logaritme van het vermogen van een spectrum, dit ziet er uit als vergelijking 6.

$$|F \{ \log (|F \{x(n)\}|^2) \}|^2 \quad (6)$$

In de situatie dat $|F \{x(n)\}|^2 = 0$ of bijna 0 is zal de berekening van vergelijking 6 een zeer grote waarde veroorzaken welke problemen in de vervolg berekeningen kan veroorzaken. In tegenstelling tot een normale cepstrum gebruikt de Mel-frequency cepstrum de Mel schaal distributie, Mel-frequency cepstrum coefficients zal in het vervolg afgekort worden tot MFCC. Deze distributie is gebaseerd op het menselijk gehoor, waardoor het betere resultaten geeft bij stemherkenning. Bij de MFCC wordt eerst het audiosignaal bewerkt met een Hamming window, waarna het spectrum wordt berekend met behulp van een Fast Fourier Transformatie. Het spectrum zal dan geschaald gaan worden met behulp van een filterbank, die doorlaat volgens de Mel schaal. De doorlaat frequenties van de filters zijn linear verdeeld tussen 0 en 1 kHz en tussen 1 kHz en 20 kHz zijn de doorlaat frequenties van de filters logaritmisches verdeeld. De responsie van de filterbank is weergegeven in figuur 1



Figuur 1: Frequentie responsie van een Mel filterbank

Het spectrum, dat het resultaat is van de toepassing van deze filterbank, kan daarna terug getransformeerd worden naar het tijddomein waaruit de coëfficiënten volgen. Om terug te transformeren naar het tijddomein wordt de discrete cosine transform methode gebruikt. Deze methode is relatief snel en heeft een hoge nauwkeurigheid. Uit de discrete cosine transform volgen 13 karakteristieken per segment. Deze karakteristieken kunnen gebruikt worden voor de herkenning van de doodskreet.

2.3 Herkennings Modellen

Er zijn verschillende methoden beschikbaar om met behulp van de korte termijn eigenschappen audioherkenning toe te passen, zoals is vermeld in Babaei, Anuar, Wahab en S. Shamsirband (2017)³ en in Vind en de Haan (2012)¹³. De meest gebruikte methoden zijn:

- Gaussian Mixture Model
- Support vector machine
- K-nearest neighbor

2.3.1 Gaussian Mixture Model

De gaussian mixture model methode zal een combinatie van twee eigenschappen van het audiosignaal tegen elkaar uit zetten. Dit kan gevisualiseerd worden door een scatter plot, waarbij de waarden van de gekozen twee korte termijn eigenschappen op de x en de y as geplaatst worden, zie de figuren in bijlage 2. Door middel van het toepassen van een gaussverdeling op de uiteenzetting van twee korte termijn eigenschappen kunnen er verschillende klassen worden gedefinieerd. Zie het werk van Skowronski en Harris(2006)¹¹ en Dhanalakshmi, Palanivel, Ramalingam (2011)¹⁴. Dit model werkt met behulp van trainingsdata. In deze data is bekend of er een doods-kreet gemaakt wordt of dat het alleen ruis bevat. Het verschil in de eigenschappen zal dan voor twee clusters van data punten zorgen, welk dan een doodskreet of een ander geluid bevat. Deze twee clusters zullen een overlap hebben en is het vaststellen van de ligging van de cluster cruciaal voor de nauwkeurigheid van dit model. Hieruit volgt ook, dat niet alle combinaties van de eigenschappen geschikt zijn voor deze methode. Want als er te veel overlap is, kunnen er geen twee verschillende clusters aan datapunten vastgesteld worden.

2.3.2 Support vector machine

Deze methode classificeert het audiosignaal met behulp van een aantal support vectors, deze methode wordt ook gebruikt in audioclassificatie door Guo en Li (2003)¹, door Temko en Nadeu (2006)⁷ en door Lu, Zhang en Jiang (2003)⁹. Deze vectoren worden gegenereerd door het trainen van een support vector machine model. Van de trainingsdata is van tevoren bekend of het een doodskreet of ruis bevat. Het voordeel van deze methode op de gaussian mixture model is, dat er voor de audio classificatie alleen maar een set vectors nodig is en er dus minder geheugen bij de classificatie nodig is.

2.3.3 K-nearest neighbor

Deze methode maakt gebruik van de euclidische afstand tussen de features van het audiosignaal en een bekende set van features. Deze bekende set van features wordt gemaakt door middel van training met behulp van datasets waarvan bekend zijn of deze bij een doodskreet of andere stalgeluiden horen. Deze methode is gebruikt door Maier, Hein en von Luxburg (2009)² en door Lu, Zhang en Jiang (2003)⁸. Bij deze methode dient dus ook getraind te worden met audiosignalen, waarbij bekend is of het doodskreten, ruis of ander geschreeuw is. Dit model zal de verschillende type signalen een klasse geven en dan bepalen bij welke klasse het ingekomen signaal het dichtst in de buurt komt.

2.4 Voorbewerkingen

Niet alles dat opgenomen wordt door de microfoons zal relevante informatie bevatten. Er zal onderzocht worden of de niet relevante geluidsbronnen onderdrukt kunnen worden, dit zou door middel van filtering uitgevoerd kunnen worden. Het beoogde doel hiervan is dat de doodskreet van een big beter te identificeren wordt. Voor de Fourier transformatie dient de audio opname geprepareerd te worden. Hiervoor wordt een Hamming window gebruikt welke het signaal op het begin en eind van het gekozen tijdsfragment onderdrukt. Hierdoor worden discontinuïteiten die zich anders in de berekeningen van de Fourier transformatie bevinden onderdrukt. Het resultaat van de Fourier transformatie zal hierdoor scherper zijn. De parameters van deze berekeningen zullen aan de hand van de eigenschappen van de doodskreet van een big bepaald moeten worden. Voor de details van deze berekeningen zie Haykin en Moher hoofdstuk 8 & 9¹⁰.

3 Onderzoeksmethode

De algemene probleemstelling en doelstelling worden beschreven en de onderzoeksopstelling wordt toegelicht.

3.1 Probleemstelling

In de intensieve varkenshouderij worden 7,2 % van de biggen doodgelegen door de zeug. In de biologische varkenshouderij is dat nog hoger: 15-25 %. Elke big, die voortijdig sterft, is een gemiste inkomst. Dus er worden opties onderzocht om deze biggen te redden. In dit onderzoek worden de doodskreten, die een big genereert, gebruikt om te detecteren of een big doodgelegen wordt. Het gebruik van doodskreten geeft al gelijk een aantal problemen. Jonge biggen maken wel ruzie met elkaar, hierbij schreeuwen ze ook. Daar dient onderscheidt gemaakt te worden. Ook is er een kans dat een big volledig onder de zeug ligt en niet instaat is om te kunnen schreeuwen, in deze situatie is het onmogelijk om de big te redden met dit alarmsysteem.

3.2 Doelstellingen

Met deze factoren inbegrepen zijn er in overleg met de opdrachtgever een aantal eisen gesteld waar het systeem uiteindelijk aan moet voldoen.

- Er moet minstens 30 % van de doodgelegen biggen geredt worden, in combinatie met het feit dat niet alle biggen gehoord kunnen worden, moet uit de meetdata 70 % van de meldingen geregistreerd worden.
- Om onnodige overlast bij het personeel te voorkomen mogen er maar 2 valse meldingen zijn per dag.
- De melding dient plaats te vinden binnen 10 seconden anders is er niet genoeg tijd om in te grijpen.

Voor dit onderzoek zijn een aantal dagen aan geluidsmetingen beschikbaar, hierbij zit ook een logboek van wanneer de biggen een doodskreet slaakte en hoe deze biggen klem zaten onder de zeug. Aan de hand van deze gegevens kunnen de doodskreten en de achtergrond geluiden geanalyseerd worden. Een gedeelte van deze data zal gebruikt worden om te analyseren en om uiteindelijke modellen te trainen en de rest wordt gebruikt om het systeem te testen.

3.3 Centrale onderzoeksvraag

Is het mogelijk om met behulp van audiosignalen uit de stal te detecteren dat een tot drie weken oude big bij de zeug in nood verkeerd?

3.3.1 Subvragen

1. Welke combinatie van audiofeatures geeft het beste resultaat in het herkennen van een doodskreet? Hier wordt onderzocht, welke combinatie van audiofeatures het mogelijk maakt om een verschil te constateren tussen de doodskreet van een big en het achtergrondgeluid, wat zich in de stal bevindt.
2. Welk herkenningsmodel werkt het beste en met welke combinatie van de audiofeatures? Er zijn verschillende combinaties van audiofeatures beschikbaar die

onderscheidt maken tussen de doodskreet van een big en het achtergrond geluid in de stal en elke combinatie levert een ander resultaat per verschillend herkenningsmodel.

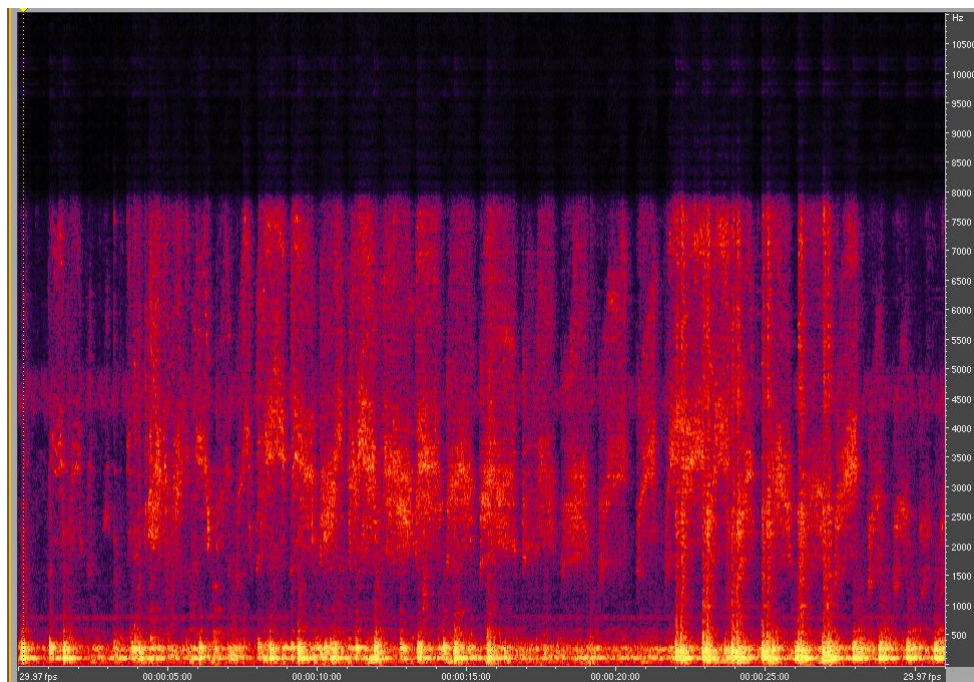
3. Is het systeem direct toepasbaar in de werkomgeving? Is het mogelijk dit systeem in de stal te gebruiken en hoeveel moet er ingesteld worden bij de installatie.

4 Signaal analyse

In dit hoofdstuk worden de beschikbare geluidsopname geanalyseerd en onderzocht wat gebruikt kan worden voor het herkenningdoeleinde. De geluidsopnames duren een half uur en zijn opgenomen met een sample frequentie van 22,5 kHz.

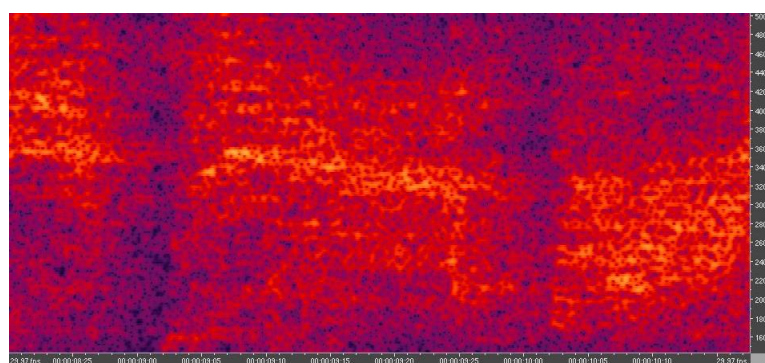
4.1 Analyse Doodskreet

Om de doodskreet te onderscheiden van andere geluiden wordt een fragment van een doodskreet eerst vergeleken met een fragment dat voornamelijk achtergrond ruis bevat. Hiermee kan onderzocht worden, welke preprocessing stappen kunnen worden uitgevoerd. Er is een spectrale analyse uitgevoerd op beide fragmenten, zodat het frequentiegebied van een doodskreet vergeleken kan worden met de achtergrondruis. De doodskreet is weergegeven in figuur 2 en de ruis is weergegeven in figuur 4.

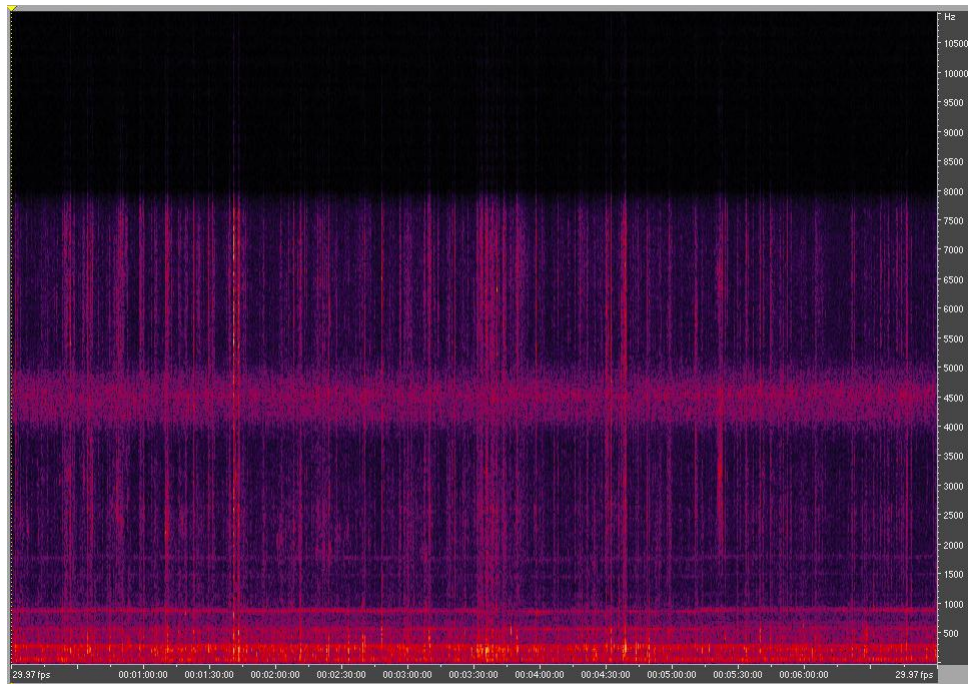


Figuur 2: Tijd/frequentie plot van een doodskreet, het lichter gebied bevat meer energie en de hoogte geeft de frequentie aan.

De gemeten frequentieband is 0-8000 Hz, hogere frequenties worden niet gemeten en zijn ook niet van toepassing voor dit onderzoek. Uit figuur 2 valt af te leiden dat



Figuur 3: Uitvergroting van de tijd/frequentie plot, geconcentreerd op het gebied dat een doodskreet bevat



Figuur 4: Tijd/frequentie plot van een audiofragment dat alleen achtergrond ruis bevat.

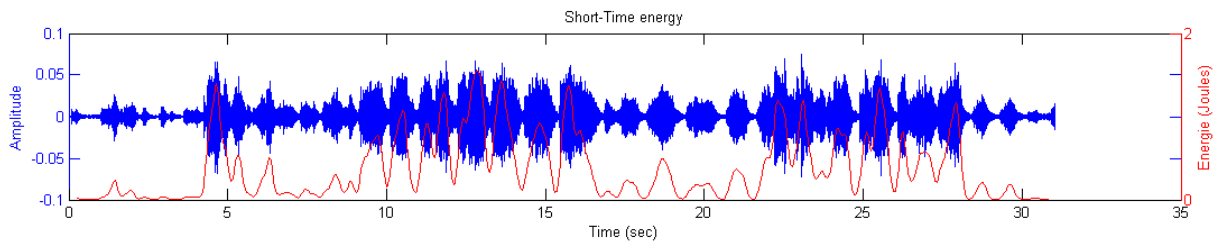
de doodskreet zich in de frequentieband van 2500-4000 Hz ligt.

Ook is het repeteren van de schreeuw zichtbaar in figuur 2, de schreeuw repeteert zich elke 500 ms. Bij de enkele schreeuw is ook de geleidelijke opbouw te zien, dit is uitgegroot in figuur 3. Deze eigenschappen kunnen gebruikt worden voor de identificatie van een doodskreet. In figuur 4 zijn de stationair achtergrondruis bronnen zichtbaar. Het ruissignaal tussen de 4000 en 5000 Hz is een ventilator, die in de stal aanwezig is. Het laag frequente ruis, 1000 Hz en minder, is achtergrond geluid dat zich in de stal bevindt. Deze twee prominent aanwezige ruisbronnen kunnen in een pre-processing stap met een bandfilter weggefilterd worden, zodat deze bronnen geen invloed hebben op het detectie algoritme. Een andere ruisbron die aanwezig is in het spectrumplot, is een algemeen verhoogd vermogen over de gehele frequentieband. De oorzaak van deze storingsbron is het klapperen van de voederbakken. Dit evenement is er niet makkelijk uit te filteren en kan bij de verdere verwerking van het signaal problemen leveren. Daarentegen is het onwaarschijnlijk dat er een big vastgeklemd zit onder een zeug wanneer deze naar de voederbak gaat.

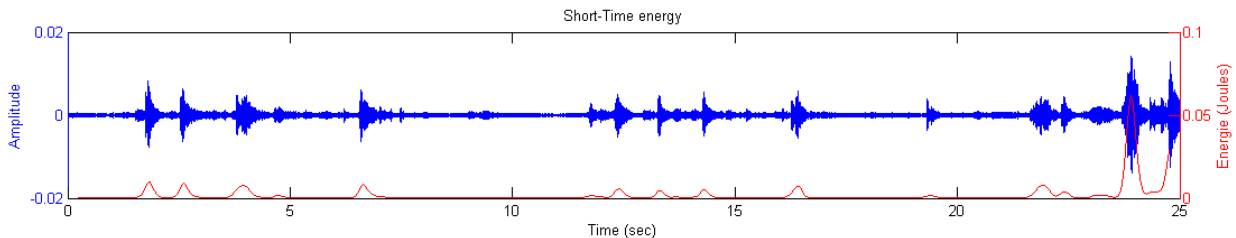
Een doodskreet varieert niet zo snel met de tijd voor een korte periode tussen de 10 en 100 ms, terwijl deze over de volledige 500 ms meer fluctueert. Dit maakt het mogelijk om met een aantal korte termijn eigenschappen een doodskreet te onderscheiden van andere geluiden. Om zulk soort eigenschappen te berekenen, dient het audiosignaal opgesplitst te worden in segmenten. Deze segmenten zijn 500 ms lang en hebben een 90 % overlap, de overlap is nodig om sneller veranderingen te detecteren en te voorkomen dat er een doodskreet gemist wordt. De korte termijn eigenschappen, onderzocht in hoofdstuk 2.2, van de doodskreet worden onderzocht of dat ze geschikt zijn voor het onderscheiden van de doodskreet en andere geluiden in de stal.

4.1.1 Short term energy

De korte termijn energie van een doodskreet wordt berekend met vergelijking 1. Het resultaat hiervan is te zien in figuur 5, ter verduidelijking is het audiosignaal ook in de figuur weergegeven. In dit figuur is blauw het audiosignaal met de schaalverdeling



Figuur 5: Een audiofragment van een doodskreet, in blauw aangegeven, met de bijbehorende korte termijns energie aangegeven in rood



Figuur 6: Een audiofragment van de achtergrond ruis, in blauw aangegeven, met de bijbehorende korte termijns energie aangegeven in rood

aan de linkerkant en rood is de korte termijn energie met de schaalverdeling aan de rechterkant. Wanneer de amplitude van het signaal stijgt, zal de korte termijn energie ook stijgen. Dit is ook zichtbaar in figuur 5. Ter vergelijking is er in figuur 6 de korte termijns energie van alleen de achtergrond ruis weergegeven. Nu lijkt er in figuur 6 op 22 en 28 seconden een piek te zitten, maar de schaalverdeling van de korte termijns energie is tien keer zo laag.

Voordelen van deze eigenschap:

Een big, die klem zit, schreeuwt met zoveel mogelijk volume en het signaal zal dan veel energie bevatten. Een big schreeuwt een periode van 500 ms, dus korte pulsen met veel energie kunnen hierdoor gefilterd worden.

Nadelen van deze eigenschap:

Als de big klem zit, zodat hij niet zo hard kan schreeuwen, zal er weinig energie in het signaal zitten en kan het misschien niet boven de achtergrondruis komen. Langdurige geluidsbronnen met veel vermogen zorgen voor storing.

4.1.2 Zero crossing rate

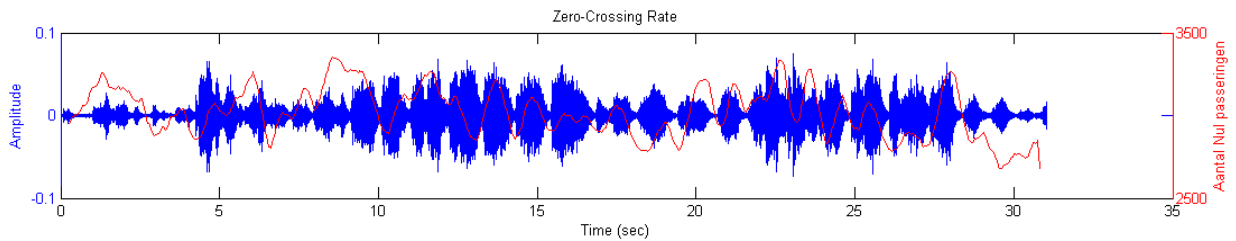
De zero crossing rate wordt berekend met behulp van vergelijking 2. Verwacht wordt dat bij ruis er een zeer hoge zero crossing rate zal zijn en op het moment dat een big gaat schreeuwen, zal de zero crossing rate moeten dalen naar de frequentie van het geschreeuw. Deze ligt zoals eerder vermeld in het gebied van 2500 tot 4000 Hz. De zero crossing rate is weergegeven in figuur 7. In dit figuur is ook het audiosignaal weergegeven, zodat zichtbaar is hoe de zero crossing rate reageert op de doodskreet. Figuur 7 heeft dezelfde opbouw als figuur 5, alleen is de korte termijns energie vervangen door de zero crossing rate. In figuur 8 is de zero crossing rate van een ruissignaal weergegeven.

Voordelen van deze eigenschap:

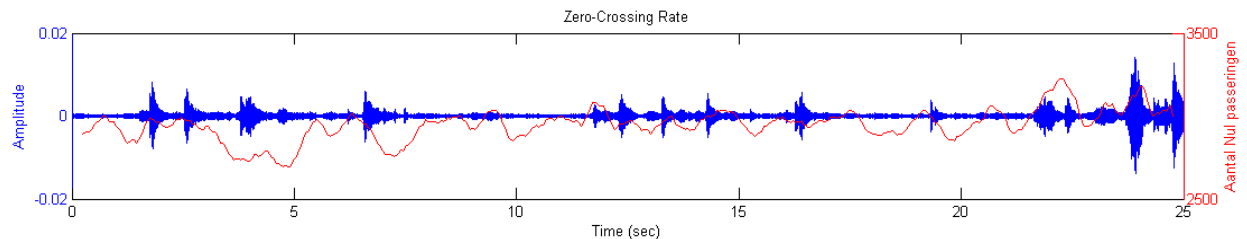
Korte pulsen hebben een hoge zero crossing rate en vallen dus weg tot de relatief lager zero crossing rate van een doodskreet. Aangezien van een doodskreet de frequentie snel verandert, is de zero crossing rate goed om de start van een doodskreet te detecteren.

Nadelen van deze eigenschap:

Het geschreeuw van een ruzie tussen biggen zit in hetzelfde frequentie spectrum als



Figuur 7: Een audiofragment van een doodskreet, aangegeven in blauw, met de bijbehorende zero crossing rate aangegeven in rood



Figuur 8: Een audiofragment van de achtergrond ruis, aangegeven in blauw, met de bijbehorende zero crossing rate aangegeven in rood

de doodskreet en kan er daar dus geen onderscheidt tussen gemaakt worden.

4.1.3 Spectral centroid

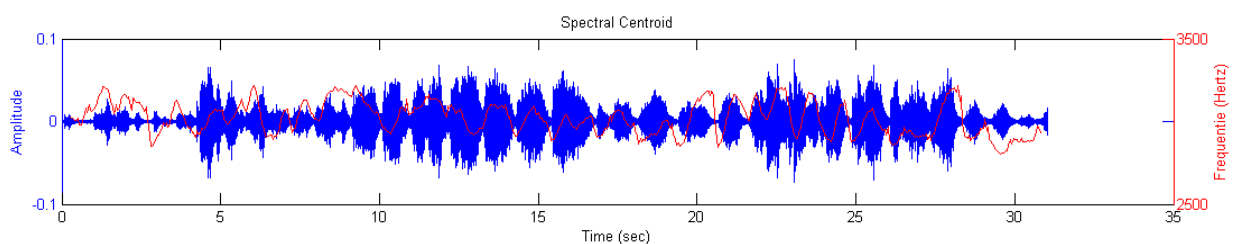
De spectral centroid wordt berekend volgens vergelijking 4. Er wordt verwacht dat de spectral centroid bij achtergrondruis in het midden van het bandfilter zal liggen en dat de spectral centroid zal stijgen wanneer een big start met schreeuwen. Uit de figuren 9 en 10 blijkt dat de spectral centroid van de ruis hoger is dan verwacht en dat deze daalt op het moment dat een big zal schreeuwen. Dit valt te verklaren, doordat het signaal gefilterd wordt om grote stationaire ruisbronnen te onderdrukken. Dit bandfilter is gespecificeerd om het signaal tussen de 2500 Hz en 4000 Hz door te laten. Om de spectral centroid van de ruis te wijzigen, dient er een ander frequentie bereik voor de filter gekozen te worden.

Voordelen van deze eigenschap:

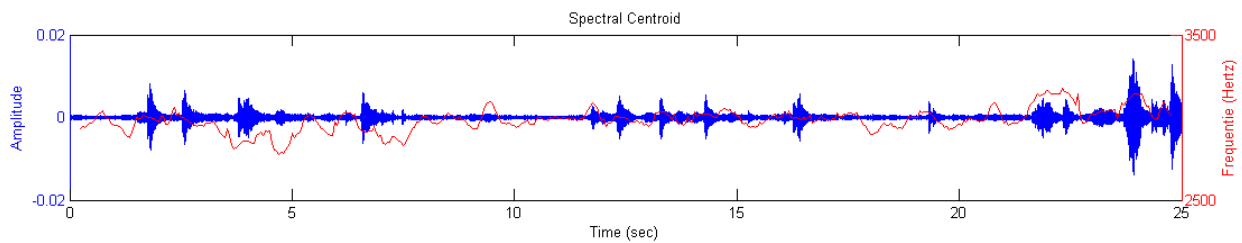
De spectral centroid is niet puur de frequentie, maar neemt ook het vermogen mee. De spectral centroid zal minder van frequentie wijzigen bij het gewone geschreeuw van biggen dan bij een big die schreeuwt voor zijn leven.

Nadelen van deze eigenschap:

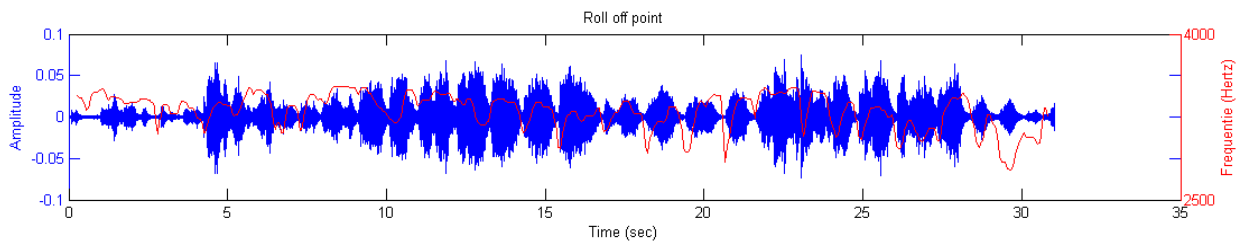
Het filterproces, dat de constante bronnen van ruis verwijdert, heeft ook invloed op de spectral centroid. Dus moet opgepast worden, dat het filter er niet voor zorgt dat de spectral centroid van de ruis overlapt wordt met die van een doodskreet.



Figuur 9: Een audiofragment van een doodskreet, aangegeven in blauw, met de bijbehorende spectral centroid aangegeven in rood



Figuur 10: Een audiofragment van achtergrond ruis, aangegeven in blauw, met de bijbehorende spectral centroid aangegeven in rood



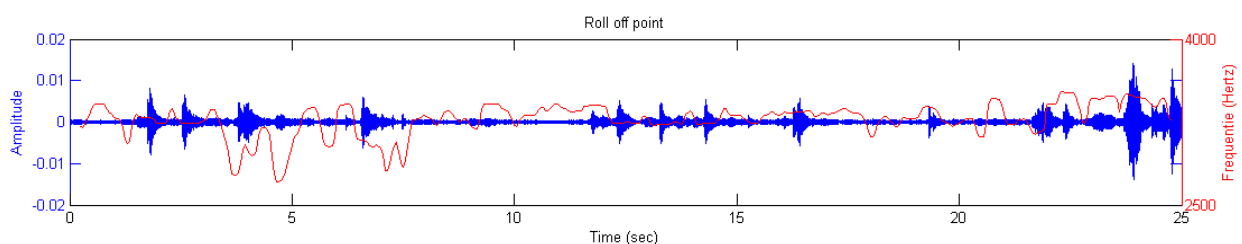
Figuur 11: Een audiofragment van een doodskreet, aangegeven in blauw, en de bijbehorende roll off point aangegeven in rood

4.1.4 Roll off point

De roll off point van een doodskreet en van een ruissignaal zijn weergegeven in figuren 11 en 12. In deze figuren is duidelijk te zien dat de roll off point bij de doodskreet de hoogst frequentie is van een schreeuwende big en bij de ruis bijna constant de hoogste frequentie is, die het onderzochte spectrum bevat. De roll off point lijkt zeer gevoelig voor ruis te zijn. Dit is zichtbaar in figuur 12 waar een paar flinke daling te zien zijn. Deze onvoorspelbaarheid zorgt ervoor dat het lastig is om onderscheid te maken tussen doodskreten en achtergrondruis.

4.1.5 Mel-frequency cepstrum coefficients (MFCC)

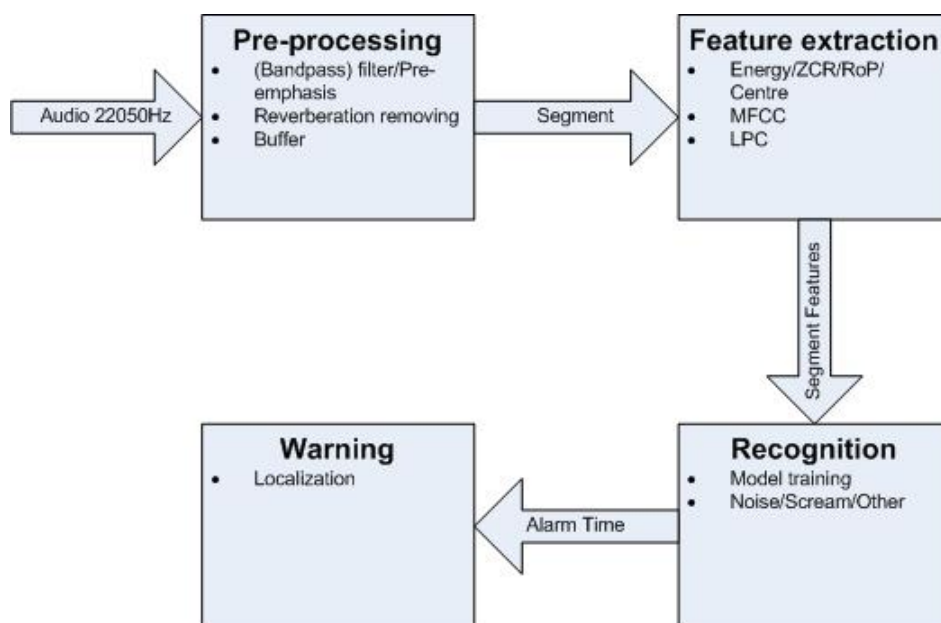
De berekening in vergelijking 6 wordt gebruikt om de karakteristieken van de doodskreet en ruis te berekenen. Deze karakteristieken van de doodskreet en ruis zijn weergegeven in bijlage 1. De relatie tussen het signaal en de karakteristiek is lastiger te zien in de figuren dan bij de andere korte termijn eigenschappen. Dit komt voornamelijk, omdat elk MFCC karakteristiek voor een kleiner frequentiebereik telt dan bijvoorbeeld de zero crossing rate, welke de frequentie van het gehele signaal schat.



Figuur 12: Een audiofragment van achtergrondruis, aangegeven in blauw, en de bijbehorende roll off point aangegeven in rood

4.2 Het systeem

Om meer inzicht te krijgen op het systeem is de werking weergegeven in figuur 13. In dit blokschema zijn een aantal stappen weergegeven, wat nodig is om een melding te genereren. Het systeem krijgt een audiosignaal van een microfoon geleverd. Op dit audiosignaal volgt dan een pre-processing. Hier worden alle onnodige en ongewenste signalen, welke buiten de band van biggen geluiden valt, weggefilterd. Dit zorgt ervoor dat stationaire stoorsignalen, die makkelijk te onderscheiden zijn van biggengescreuw, niet in de verdere processen zitten. Ook wordt het signaal gebufferd, zodat er makkelijker meegewerkt kan worden in de volgende proces stappen. Uit het segment dat hieruit volgt, zal de gekozen eigenschappen voor het model berekend worden. Deze eigenschappen zijn verder verklaard in hoofdstuk 4.1. Met deze eigenschappen zal dan een model bepalen of er een doodskreet is of niet. Als er een doodskreet is, zal het systeem een tijdstip uitgeven van wanneer deze kreet gestart is. Daarna zal nog bepaald moeten worden in welk hok de biggen aan het schreeuwen zijn, maar dat valt buiten de doelstelling van dit project.



Figuur 13: Systematische weergave van het alarmsysteem

5 Model Training

Nu verschillende detectie modellen onderzocht zijn en de karakteristieke eigenschappen van de biggen schreeuwen bekend zijn, worden de modellen getrained met de eigenschappen van de biggen schreeuw.

5.1 Feature selectie

Eerst zal er onderzocht moeten worden, welke features van de doodskreten van de biggen het meest geschikt zijn om de herkenning modellen mee te trainen. Hiervoor zullen de features tegen elkaar uit gezet worden en onderzocht worden of er voldoende verschil is met ruis om de situatie te herkennen.

De vijf geluids features:

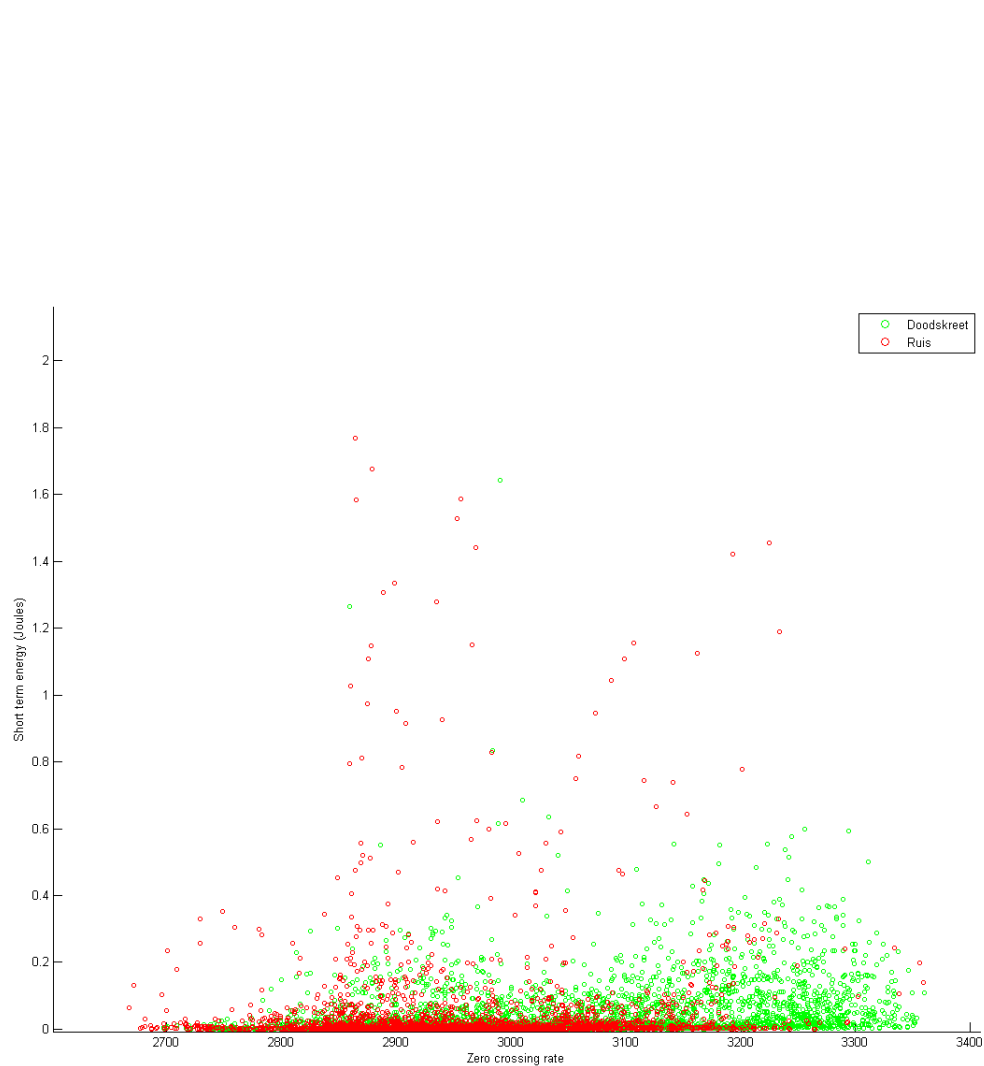
- Short term energy
- Zero crossing rate
- Spectral centroid
- Roll off point
- Mel-frequency cepstrum coefficients

Twee geluids features worden tegen elkaar uitgezet en dan wordt onderzocht of er voldoende verschil is tussen een doodskreet en achtergrond geluiden in de stal om onderscheid te maken. En er wordt onderzocht welke combinatie tussen de geluids features het meeste verschil heeft.

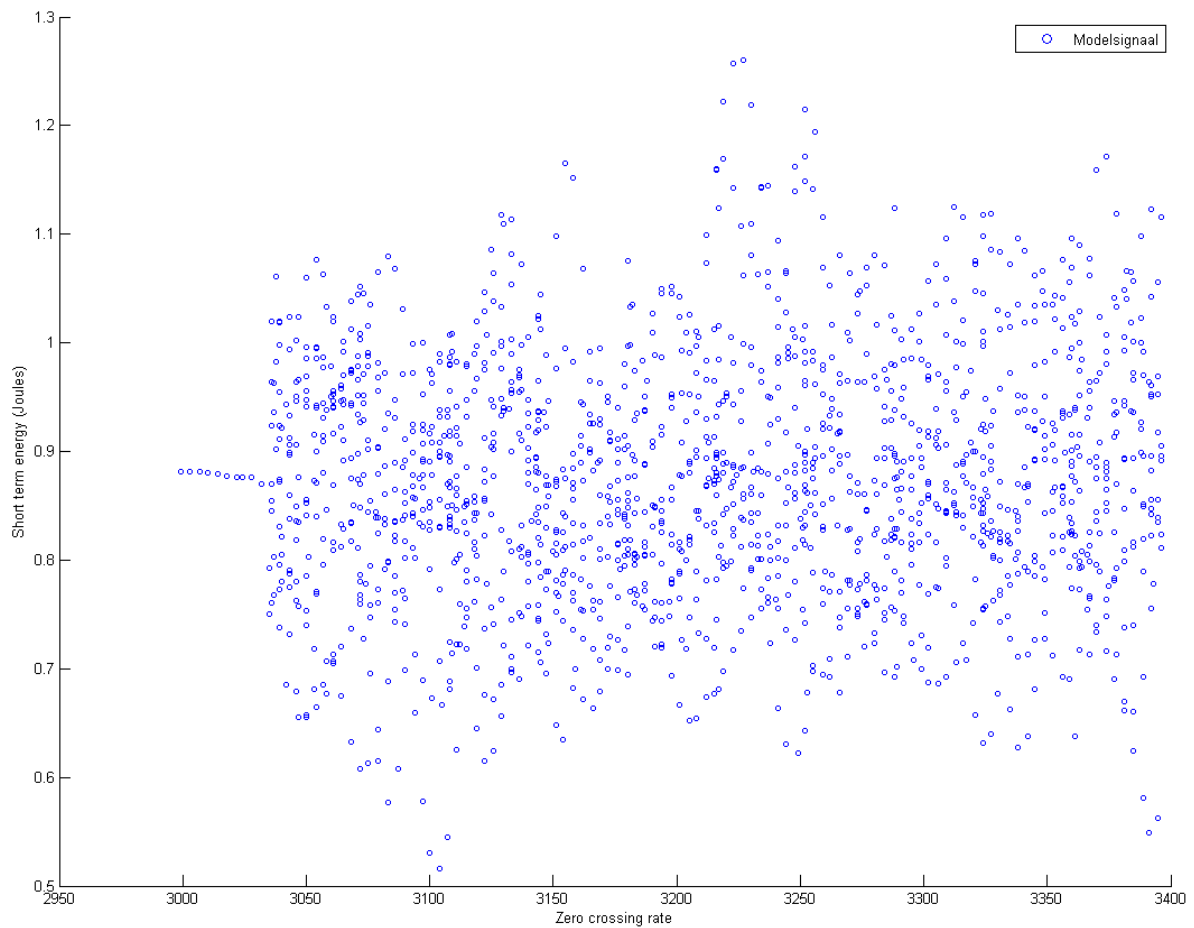
In figuur 14 wordt de short term energy, de horizontale as, en de zero crossing rate, de verticale as, tegen elkaar uitgezet. In de figuur zijn twee klassen weergegeven, de rode klasse is het ruissignaal en de groene klasse is de doodskreet van een big. Er is tussen deze features nog steeds overlap, maar niet zoveel als bij de andere onderzochte combinaties. Voor de duidelijkheid is het ruissignaal in rood weergegeven en de doodskreet van een big in groen. Dit is besloten om duidelijk onderscheid te observeren tussen een noodsituatie en de achtergrondruis. Om te kunnen beslissen welke situatie zich voordoet dient er een gebied te zijn waarbij er geen overlap is tussen het ruissignaal en de doodskreet. De resultaten van alle onderzochte combinaties is te vinden in bijlage 2.

5.1.1 Feature modellering

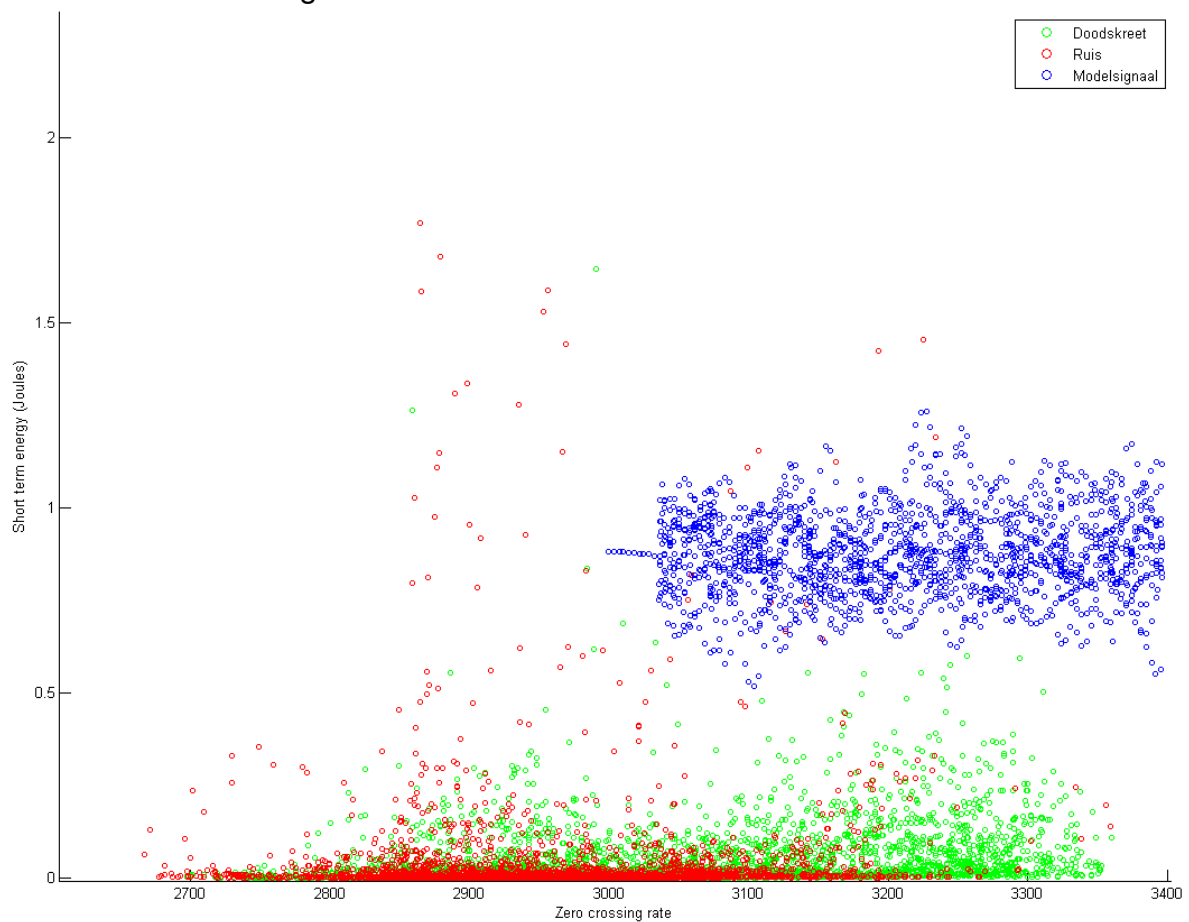
Er wordt ook een model gemaakt van een doodskreet van een big, waar dan ook de audiofeatures van berekend kunnen worden. Aan de hand van dit model kan onderzocht worden hoeveel invloed ruis heeft op het resultaat van de model trainingen en de uiteindelijke detectie. Dit model zal worden opgebouwd aan de hand van de analyse van de doodskreet in hoofdstuk 2. De features van dit model zijn weergegeven in figuur 15 en in figuur 16 zijn de features weergegeven met de features van de ruis en de doodskreten als referentie. Zoals te zien in de figuren 15 en 16 bevat het model frequenties tussen de 3050 en 3400 Hertz. Dit is het belangrijkste gebied waar de doodskreten van de biggen starten. Er is voor gekozen om de amplitude van het modelsignaal te variëren, zodat de features een groter gebied omvatten dat voornamelijk overlap heeft met de features van doodskreten.



Figuur 14: Short term energy van een doodskreet, weergegeven in groen, en van achtergrondruis, weergegeven in rood, uitgezet tegen de zero crossing rate van dezelfde signalen



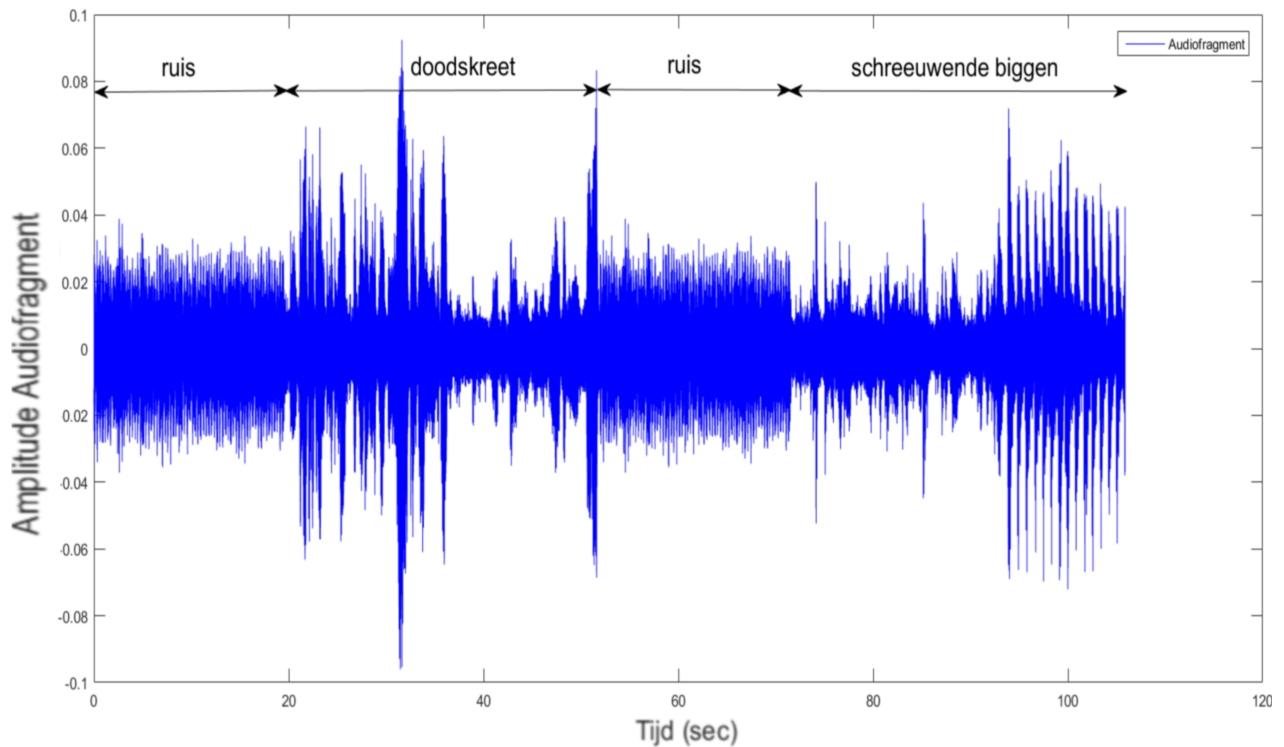
Figuur 15: Short term energy van het model signaal uitgezet tegen de Zero crossing rate van het model signaal



Figuur 16: Figuur 14 en figuur 15 samengevoegd

5.1.2 Testsignaal

Om de methoden te testen wordt er een testsignaal gemaakt. Dit signaal is opgebouwd uit verschillende audiofragmenten. Het testsignaal dient de belangrijke geluiden die in de stal voorkomen te bevatten. Het eerste gebruikte audiofragment bevat daarom alleen ruis dat opgenomen is in de stal, dit audiofragment is ongeveer 20 seconden. Het tweede audiofragment bevat een doodskreet van een big die in gevaar is, dit audiofragment is ongeveer 32 seconden. Hierna volgt het eerst gebruikte audiofragment weer, dit zorgt voor een rustige overgang naar het laatste audiofragment. Het laatste audiofragment bevat de geluiden van een paar schreeuwende biggen die niet in gevaar zijn. Het resulterend test signaal is weergegeven in figuur 17 en zal gebruikt worden om de detectie methoden te analyseren.



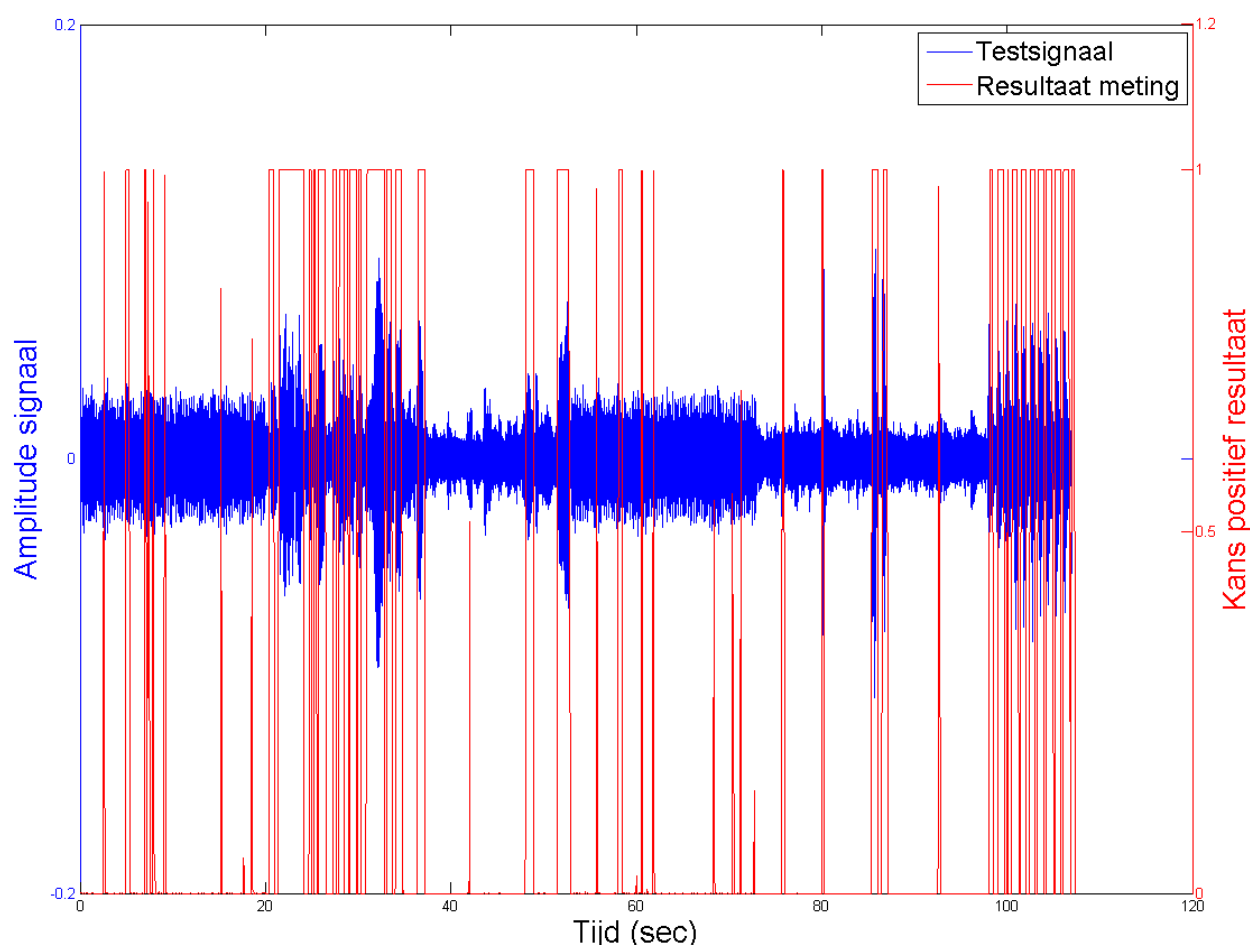
Figuur 17: Testsignaal opgebouwd met audiofragmenten van achtergrondruis, een doodskreet en ruziënde biggen

5.2 Gaussian mixture model

Hier zal onderzocht worden of er voldoende verschil is tussen een doodskreet en andere geluiden in de features van de audiofragmenten.

5.2.1 Model training

Het model wordt getraind met de feature geselecteerd in hoofdstuk 4 en daarna wordt het model getest met het testsignaal dat weergegeven is in figuur 17. Het resultaat is te zien in figuur 18. De doodskreet bevindt zich in het centrum van het audiofragment, de amplitude van het signaal fluctueert hier veel meer. Wanneer het rode signaal 1 aangeeft, wordt volgens het gaussian mixture model een doodskreet gedetecteerd. Dit gebeurt ook in het ruisgebied. Hieruit valt te concluderen dat het model te veel false positives heeft. Er zal verder onderzocht moeten worden hoe deze hoeveelheid false positives gereduceerd kunnen worden.

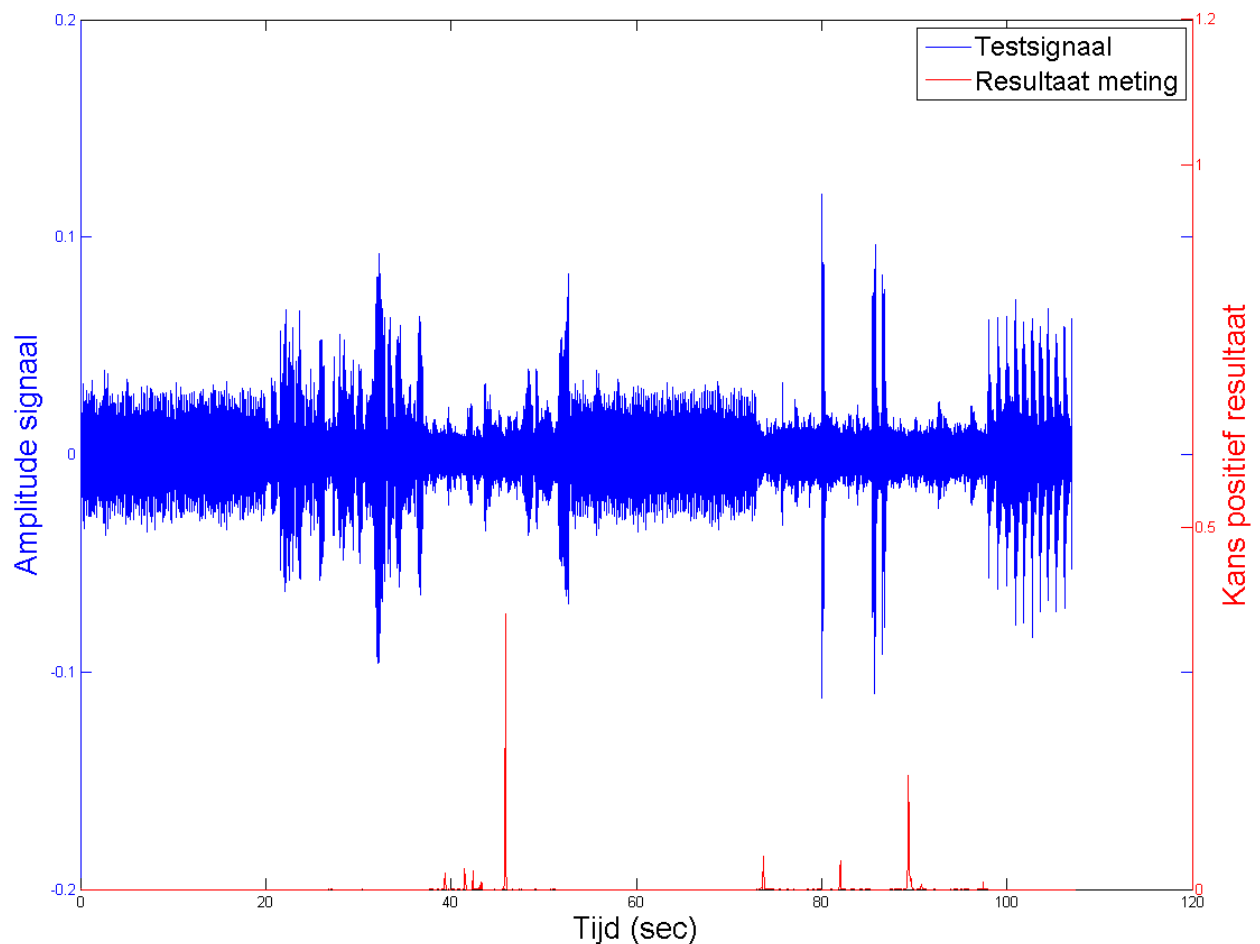


Figuur 18: Testsignaal, aangegeven in blauw, met het resultaat van een gaussian mixture model meting, aangegeven in rood, met de short term energy en de spectral centroid als features

5.2.2 Model training met behulp van gemodelleerd signaal

Om te onderzoeken of er verbetering te vinden is door een model te maken van een doodskreet, wordt daarmee nu het gaussian mixture model getraind. Uiteraard worden dezelfde features geselecteerd als bij de standaard model training. Het getrainde gaussian mixture model wordt getest met hetzelfde audiofragment en het resultaat hiervan is te zien in figuur 19. Bij vergelijking van de resultaten, zichtbaar in de figuren

18 en 19, valt op dat door het gebruik van een gemodelleerde doodskreet er geen false positives zijn. Ook geeft het gaussian mixture model nauwelijks de doodskreet aan, hieruit blijkt dat het modelsignaal te ongevoelig is ingesteld voor het gaussian mixture model. Deze gevoeligheid zou verhoogd kunnen worden door het modelsignaal hoger frequent te maken en het vermogen van het modelsignaal te reduceren, zie figuur 16. Het modelsignaal zal dan meer overlap krijgen met het doodskreet gebied, maar het zal ook dichterbij het ruisgebied komen.



Figuur 19: Testsignaal, aangegeven in blauw, met het resultaat van een gaussian mixture model meting, aangegeven in rood, met de short term energy en de spectral centroid van het modelsignaal

5.3 Support vector machine

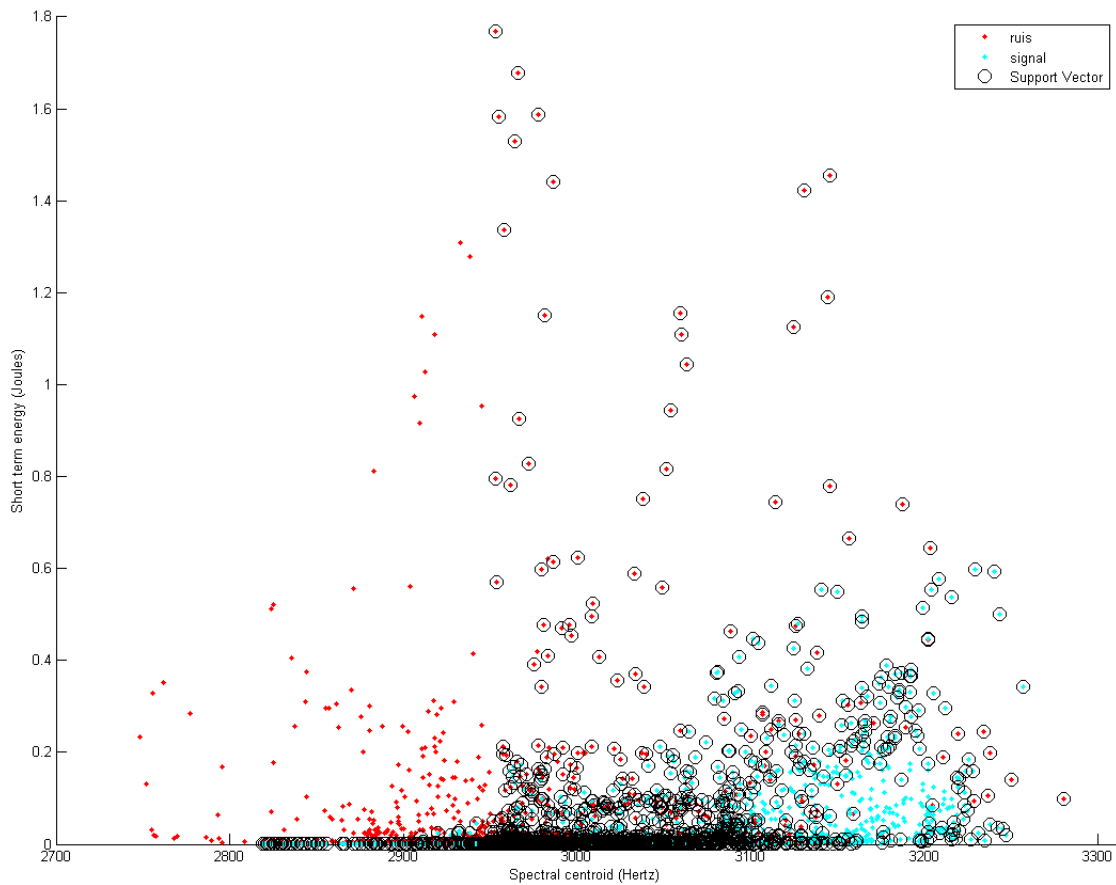
Deze methode werkt eveneens als het gaussian mixture model door het tegen elkaar uit zetten van de features van de audiofragmenten die gebruikt worden bij het onderzoek.

5.3.1 Model training

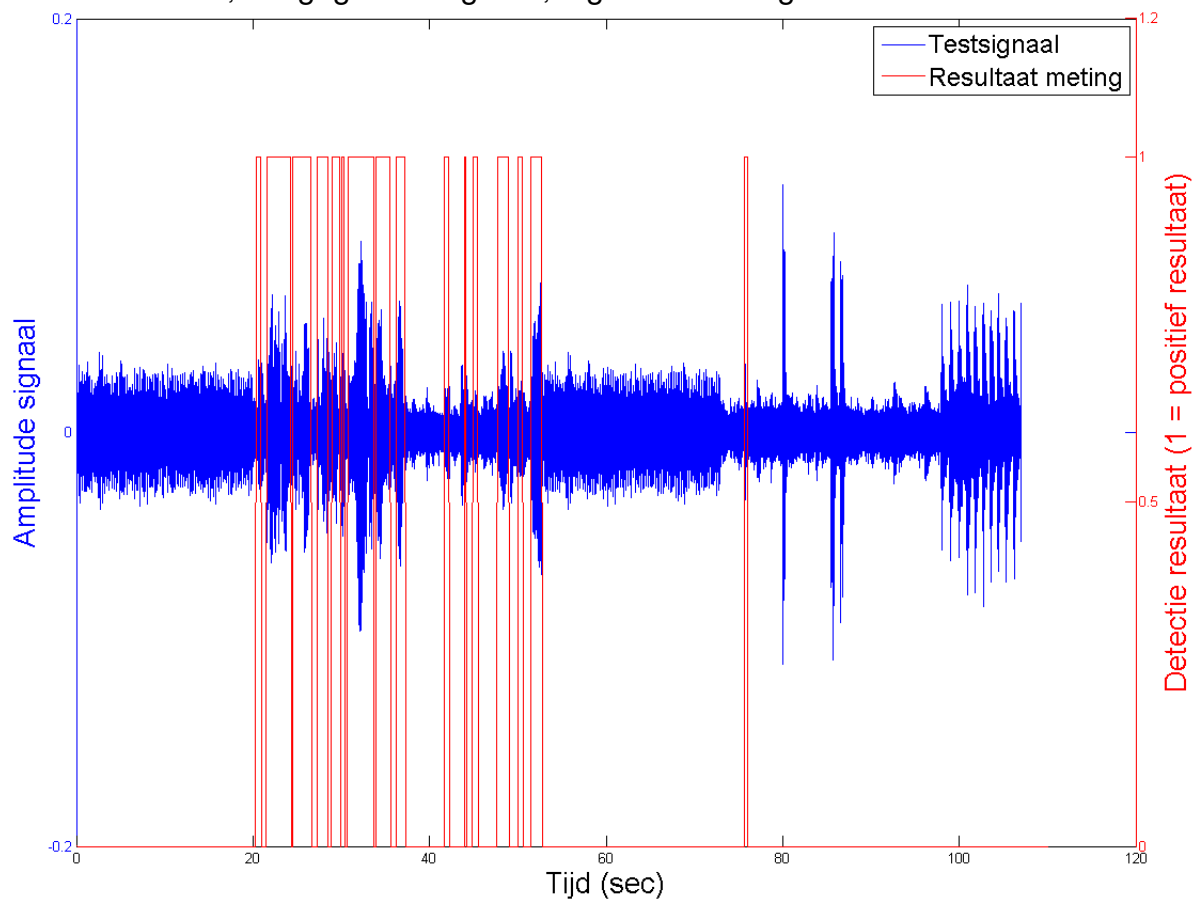
De support vector machine wordt getraind met de geselecteerde features van hoofdstuk 4. Het resultaat hiervan is te zien in figuur 20. In figuur 20 is de spectral centroid tegen de short term energy uitgezet. Het valt op dat in dit figuur een significante hoeveelheid support vectors bevinden in een laag energie gebied tussen de 2800 en 3150 Hertz. Dit komt waarschijnlijk doordat er veel overlap in het gebied is tussen het ruis-signaal en de doodskreten. Maar verrassend is dat de support vector machine de data tussen de 2900 en 3000 Hertz en de energie tussen de 0,2 en 1,8 klassificeert als ruis, terwijl in het gebied zich voornamelijk doodskreten bevinden. Het is dan ook handig om bij de trainingsdata zuivere data te gebruiken voor de doodskreten, aangezien de vervuiling van de ruis de klassificatie negatief beïnvloed. Het getrainde model wordt op het in hoofdstuk 5.2 geconstrueerde audiofragment getest. Het resultaat hiervan is te zien in figuur 21. Het detectie resultaat omvat de doodskreet, er is duidelijk zichtbaar dat de support vector machine niet detecteert wanneer de big inademt tussen de schreeuwen en dat kracht van de schreeuw ook afneemt met de tijd. Ook valt op dat er geen false positives zijn bij het ruisfragment.

5.3.2 Model training met behulp van gemodelleerd signaal

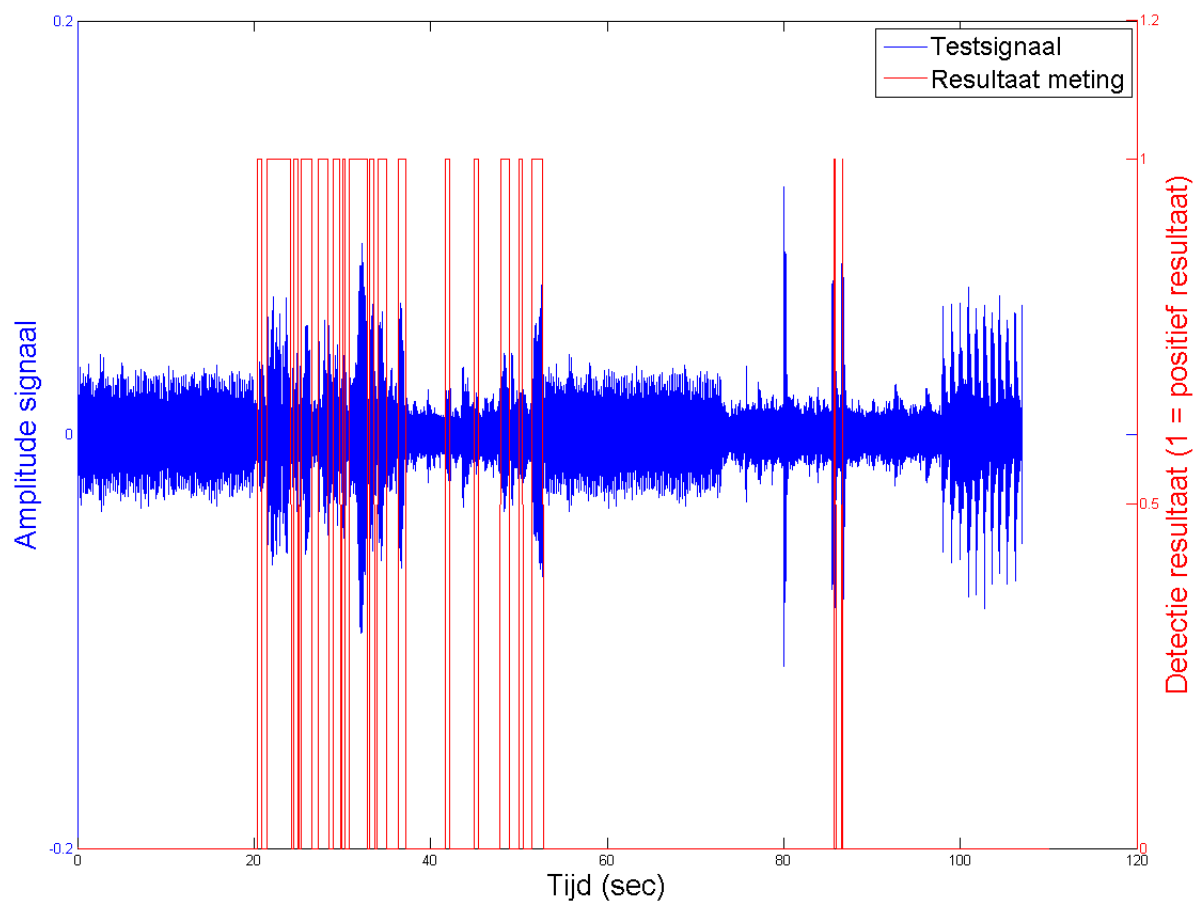
De support vector machine wordt getraind met de features van het gemodelleerd signaal. Dezelfde features worden gekozen als bij de training met de audiofragmenten. Het resulterend model wordt gebruikt om hetzelfde geconstrueerde audiofragment te klassificeren. Het resultaat van deze test is te zien in figuur 22. In dit figuur is zichtbaar dat er geen false positives zijn en dat er minder segmenten van de doodskreet gedetecteerd wordt. Dit kan resulteren dat het huidige model ervoor zorgt dat er minder false positives zijn in het geval dat er veel lawaai in de stal is.



Figuur 20: Support vector machine resultaten, aangegeven door omcirkeling, met de spectral centroid en de shortterm energy van achtergrondruis, aangegeven in rood, en een doodscreet, aangegeven in groen, tegen elkaar uitgezet



Figuur 21: Testsignaal, aangegeven in blauw, met het resultaat van de support vector machine, aangegeven in rood, met de zero crossing rate en de short term energy als features



Figuur 22: Testsignaal, aangegeven in blauw, met het resultaat van support vector machine, aangegeven in rood, met zero crossing rate en short term energy van het modelsignaal als features

5.4 K-nearest neighbour

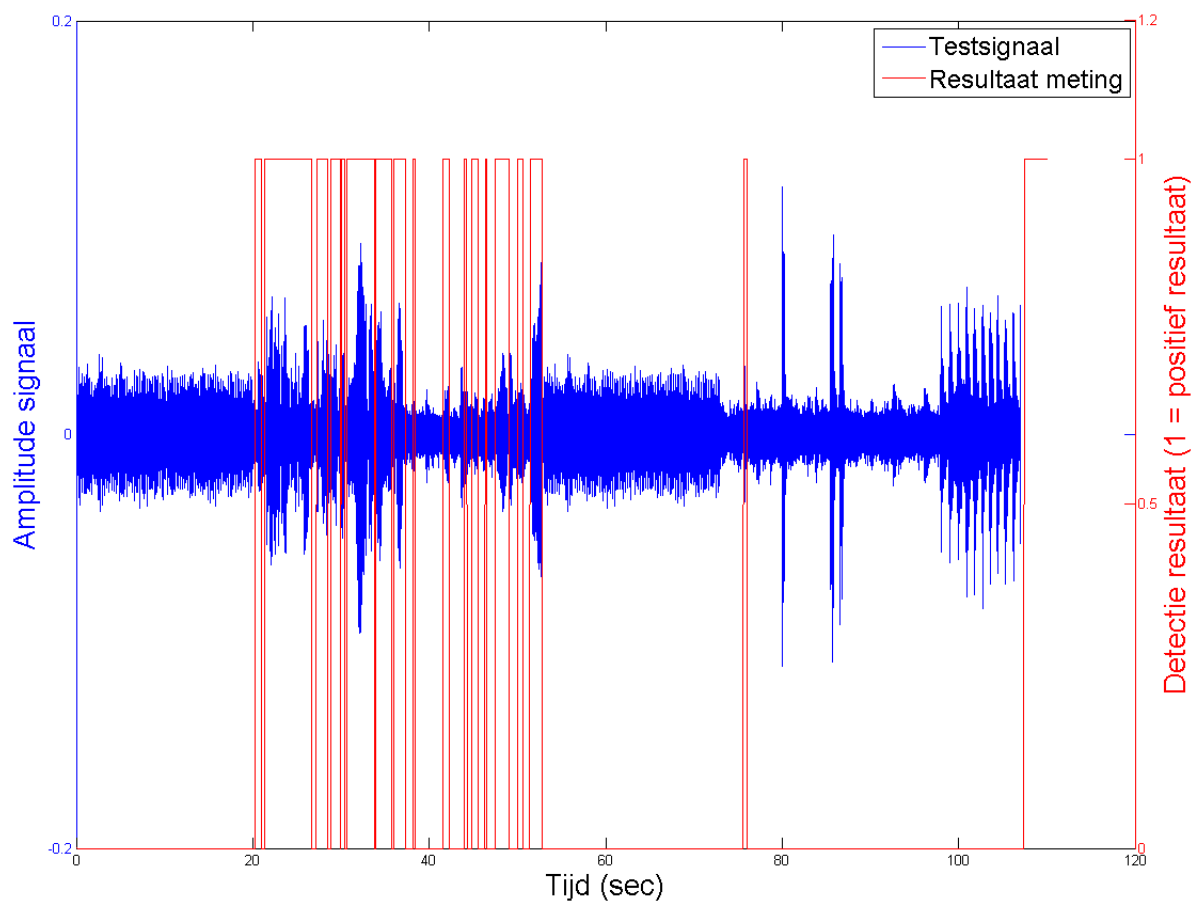
De K-nearest neighbour methode gebruikt de euclidische afstand tussen de locatie van de gemeten features en de dichtsbijzijnde gekalibreerde features om te beslissen bij welke klasse het signaal hoort.

5.4.1 Model training

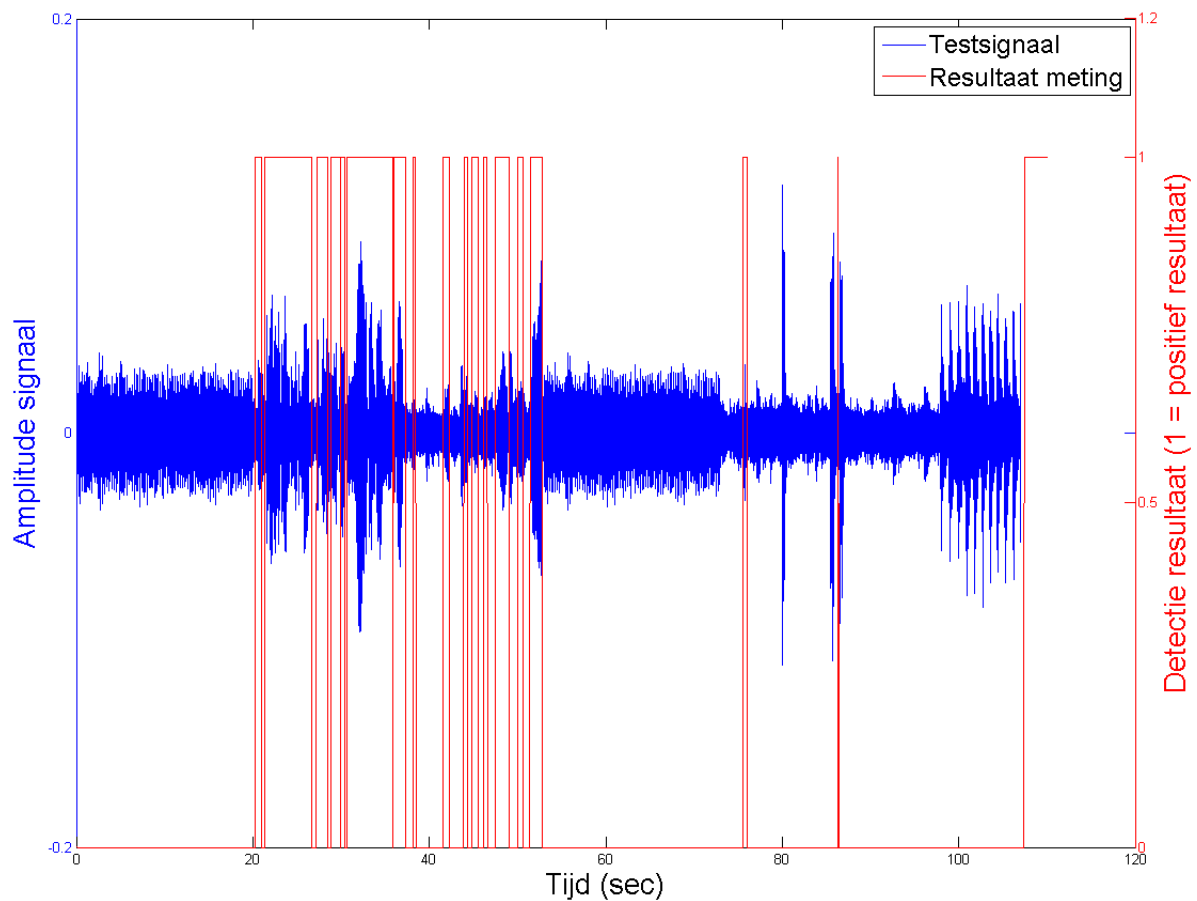
Een voordeel van de K-nearest neighbour methode is dat er meerdere van de beschikbare features gebruikt kunnen worden in plaats van maar twee features bij de gaussian mixture model en de support vector machine. Er moet dus een selectie uit de features gemaakt worden welke alleen positief bijdragen. Omdat de dataset van de features als coördinaten gebruikt worden bij deze methode, is een visuele representatie niet overzichtelijk indien er meer dan twee feature datasets gebruikt wordt. Hierdoor zal de effectiviteit van de gekozen features volgen uit de verwerking van de getrainde data. De getrainde data wordt daarom getest op een klein stukje signaal dat zowel ruis als doodskreten bevat. Dit is hetzelfde audiofragment dat gebruikt wordt in 5.2. Het eerste resultaat is te zien in figuur 23. De aanwezige doodskreet wordt goed omvat door het detectie resultaat, maar dit resultaat is gegenereerd door met 1 vergelijkingspunt de doodskreet te detecteren. Dit resulteert dat het model niet robuust is en het dus ook een positief resultaat geeft op ander geluid in de stal. Dit kan verbeterd worden door het aantal vergelijkingspunten te verhogen, zoals te zien is in figuur 24. In figuur 24 wordt de klasse bepaald door de gemiddelde euclidische afstand met de 10 dichtstbijzijnde punten van het K-nearest neighbour model. Er is een subtiel verschil zichtbaar met de detectie van de doodskreet, wat er voor kan zorgen dat dit model minder gevoelig is voor andere geluiden. In figuur 25 zijn 100 vergelijkingspunten gebruikt. Het detectie model geeft nog steeds de doodskreet goed weer. Het toevoegen van nog meer vergelijkingspunten heeft een minimum aan effect en verhoogd de benodigde rekenkracht voor het bepalen van een klasse.

5.4.2 Model training met gemodelleerd signaal

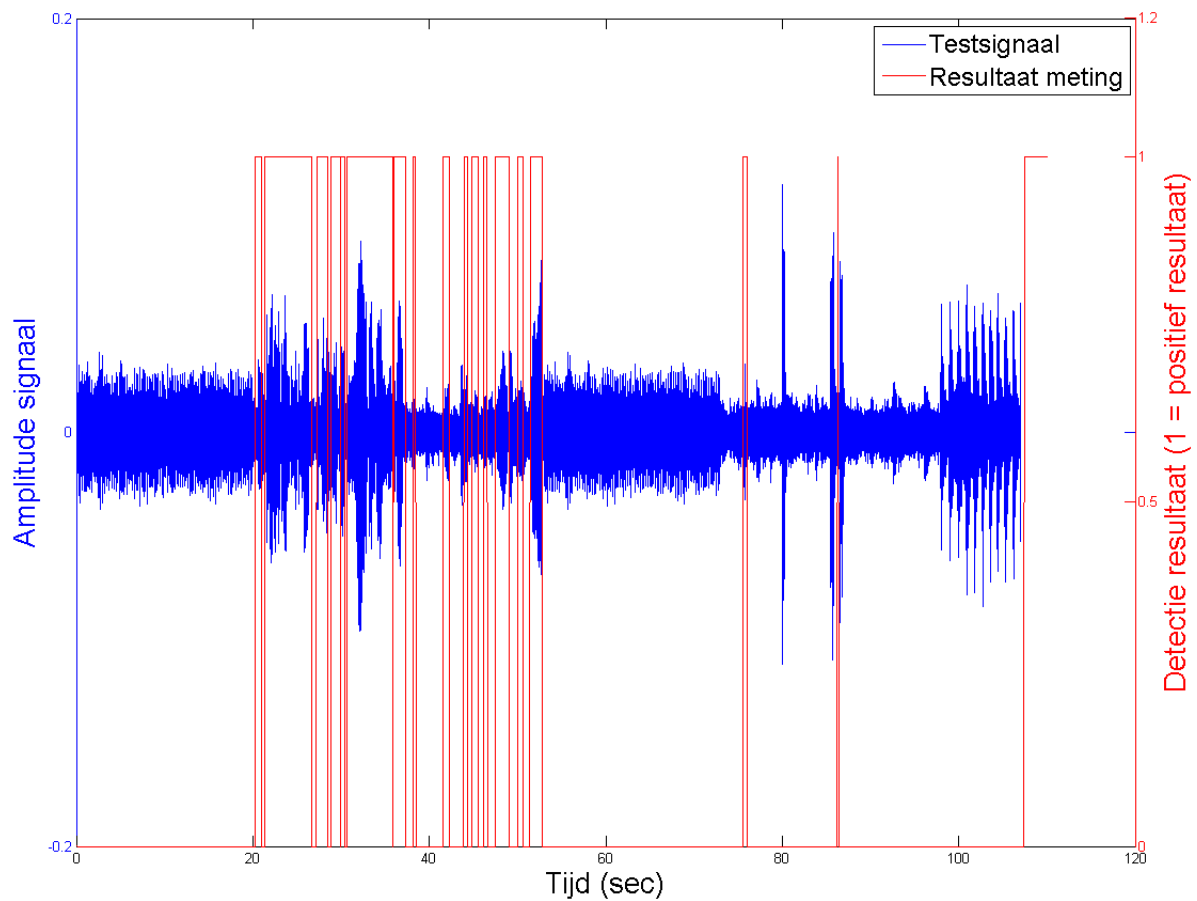
De K-nearest neighbour model wordt ook getraind met het gemodelleerd signaal. De training gebeurt met dezelfde feature selectie als bij de gewone model training. Ook wordt dezelfde hoeveelheid vergelijkingspunten gebruikt. Dit wordt dan na de training met hetzelfde datafragment getest om te vergelijken of het gebruik van een gemodelleerd signaal leidt tot verbeteringen. Het resultaat is te zien in figuur 26. In dit figuur wordt de doodskreet gedetecteert en in het ruisfragment bevindt zich geen false positive.



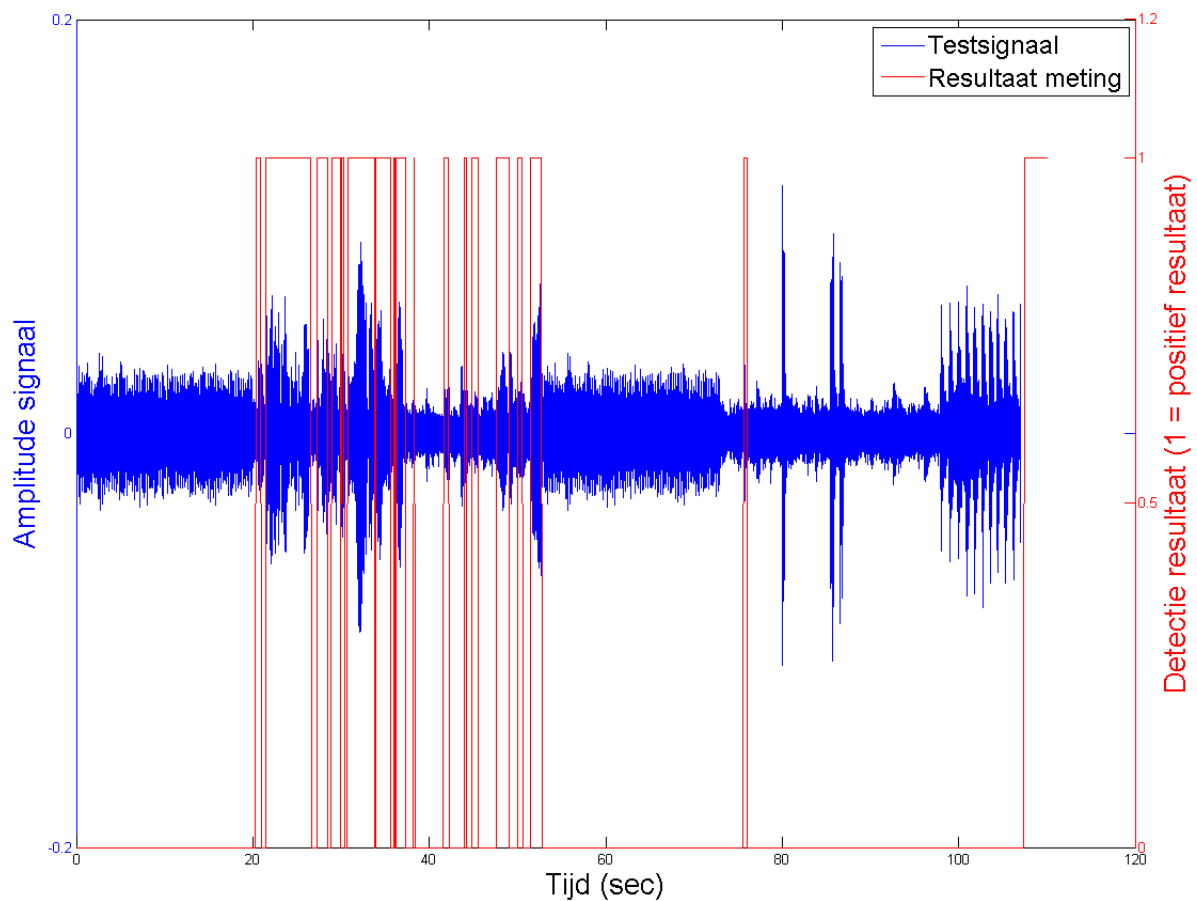
Figuur 23: Testsignaal, aangegeven in blauw, met resultaat van K-nearest neighbour, aangegeven in rood, met spectral centroid, short term energy, zero crossing rate en 1 vergelijkingspunt



Figuur 24: Testsignaal, aangegeven in blauw, met resultaat van K-nearest neighbour, aangegeven in rood, met spectral centroid, short term energy, zero crossing rate en 10 vergelijkingspunten



Figuur 25: Testsignaal, aangegeven in blauw, met resultaat van K-nearest neighbour, aangegeven in rood, met spectral centroid, short term energy, zero crossing rate en 100 vergelijkingspunten



Figuur 26: Testsignaal, aangegeven in blauw, met resultaat van K-nearest neighbour, aangegeven in rood, met spectral centroid, short term energy, zero crossing rate van het modelsignaal en 100 vergelijkingspunten

6 Conclusie

De resultaten van de verschillende herkenningsmodellen worden vergeleken en waaruit een uiteindelijke conclusie getrokken kan worden.

6.1 Resultaten tabel

De resultaten van de metingen zijn te zien in tabel 1.

Deze metingen zijn gemaakt van een 30 minuten durende geluidopname welke drie noodsituaties bevat.

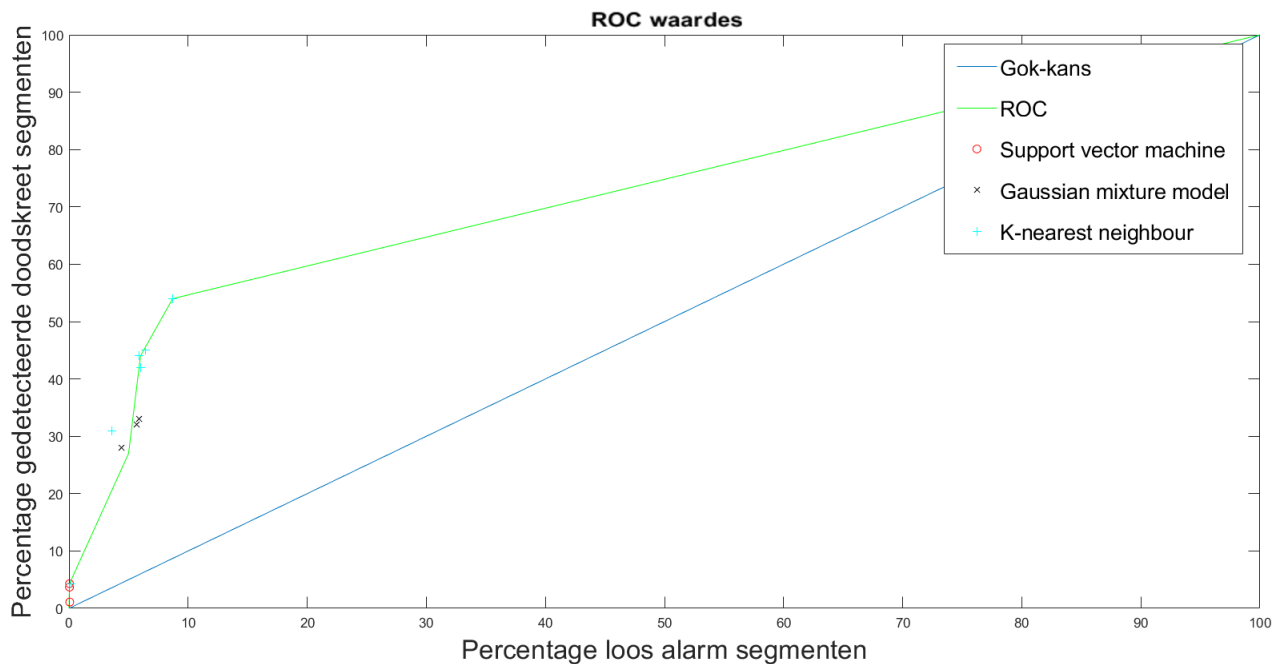
Deze geluidopname word verdeeld in 34414 segmenten waarvan 1315 segmenten noodsituaties zijn. De segmenten zijn 500 ms groot en hebben een 90% overlap met het voorgaande segment.

Tabel 1: Resultaten van verschillende methodes en hun combinaties van audiofeatures

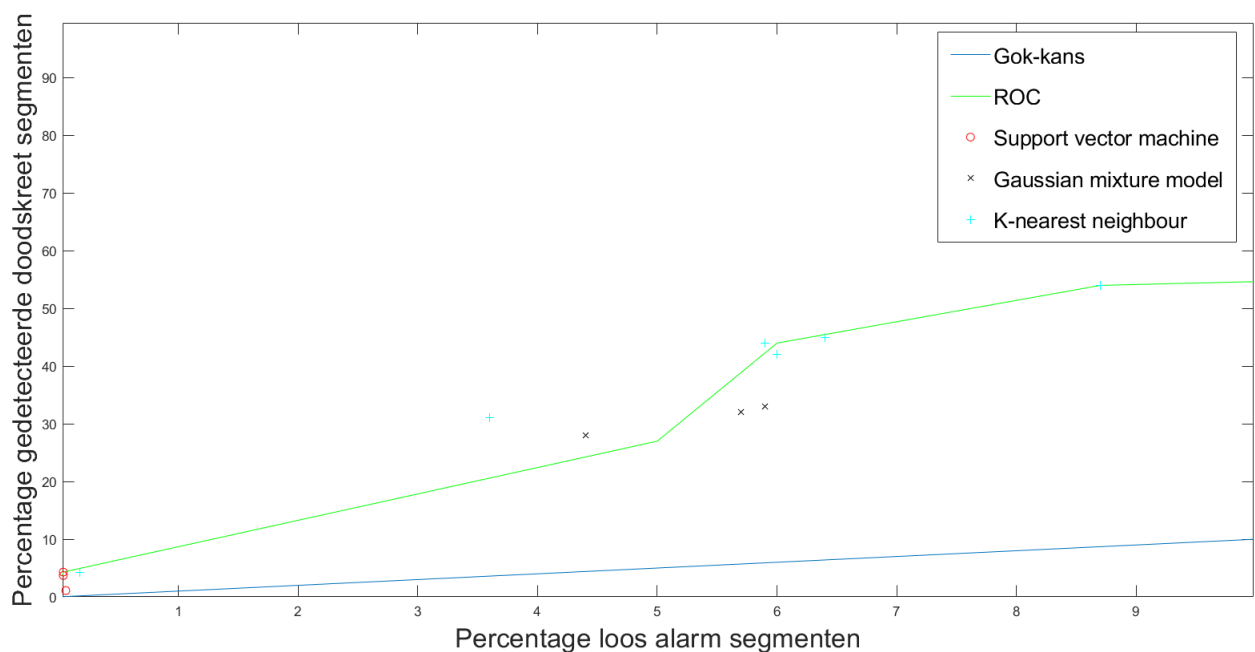
Methode	Audiofeatures	True positives	False positives	noodsituaties gedetecteerd
Gaussian mixture model	Short term energy, zero crossing rate	429	1952	3
	Short term energy, MFCC	370	1472	3
	Short term energy, spectral centroid	426	188 7	3
K-nearest neighbor	Short term energy, MFCC	55	60	2
	Short term energy, spectral centriod	597	2122	3
	Short term energy, zero crossing rate, MFCC	713	2887	3
	Short term energy, spectral centroid, MFCC	579	1942	3
	Short term energy, spectral centroid, zero crossing rate	406	1195	3
	Short term energy, spectral centroid, zero crossing rate, MFCC	547	1989	3
Support vector machine	Short term energy, zero crossing rate	57	13	3
	Short term energy, MFCC	14	20	2
	Short term energy, Spectral centroid	49	13	3

6.2 Receiver operating characteristic figuur

Receiver operating characteristic figuur, weergegeven in figuur 27 gemaakt aan de hand van de gedetecteerde doodskreten en de loze alarmen van de onderzochte methoden. Het percentage doodskreet segmenten is berekend door het percentage te nemen van het aantal segmenten van een noodsituatie die gedetecteerd zijn ten opzichte van het totale aantal noodsituatie segmenten. Het percentage loos alarm segmenten is berekend door het percentage te nemen van het aantal segmenten die foutief een noodsituatie aangeven ten opzichte van het totale aantal segmenten waar geen noodsituatie voordoet. In figuur 28 zijn de meetpunten voor het receiver operating characteristic figuur uitvergroot, zodat de verdeling van de meetpunten duidelijker is.



Figuur 27: Receiver operating characteristic van de verschillende methoden



Figuur 28: Uitvergroting meetpunten receiver operating characteristic figuur

6.3 Conclusie

Uit de resultaten tabel kan het herkenningsmodel gekozen worden welke de minste false positives heeft en nog steeds in staat is de doodskreet van de biggen te detecteren. Aan de hand van dit resultaat kan beoordeeld worden of de doelstelling van dit onderzoek gehaald is. Dus dat 70 % van de doodskreten gedetecteerd is en dat het aantal valse meldingen onder de 2 per dag zit. Als laatste is het bij elke herkenningsmodel gelukt om de berekentijd onder de 10 seconden te houden. Het langst, duurt het berekenen van de geluidsfeatures. Er zit namelijk een buffering in dit proces, wat er voor zorgt dat dit onderdeel altijd 0.5 seconden duurt. Daar komt de tijd dat het kost om de features te berekenen bovenop, maar met een moderne computer is deze hoeveelheid tijd klein ten opzichte van de buffertijd.

Tijdens het onderzoek blijkt dat de support vector machine met de short term energy en de zero crossing rate het beste resultaat heeft. Deze combinatie is in staat om de doodskreet te detecteren, maar voldoet niet aan de eis van maximaal 2 valse meldingen per dag. Het is ook gebleken dat vooral de combinatie van de short term energy en de zero crossing rate voor de meest herkenbare resultaten zorgde. Dit komt voornamelijk doordat de frequentie van de schreeuwende big, die in nood is, hoger ligt dan de meeste achtergrond geluiden en dat de harde storingsbronnen ook in een ander frequentie bereik zitten. Er zijn helaas nog wel storingsbronnen die zowel een hoge intensiteit hebben als in het frequentie bereik vallen van een doodskreet. In de huidige status van het systeem is dit niet direct toe te passen in een varkensstal. Het huidige algoritme beschikt nog niet over een calibratie methode voor de microfoons. Hierdoor kan er geen rekening gehouden worden met de verschillen in instelvermogen van een microfoon. Het instelvermogen beïnvloedt de short term energy, wat ervoor kan zorgen dat de detectiemethodes niet in staat zijn om onderscheid te maken tussen doodskreten en ander lawaai. Ook is de code geschreven om te werken met audio opnames van een half uur, terwijl het systeem uiteindelijk moet werken met een continu audiosignaal dat door de microfoon wordt opgenomen. Het huidige systeem werkt momenteel met een half uur aan opgenomen audiofragmenten en bepaalt achteraf of en wanneer er doodskreten in het audiofragment plaatsvinden.

De uiteindelijke conclusie is: het is mogelijk om met behulp van audiosignalen uit de stal te detecteren dat een één tot drie weken oude big bij de zeug in nood verkeerd. Het huidige systeem voldoet alleen niet aan alle eisen en daarvoor dient dan ook verder onderzoek te volgen. Het systeem voldoet niet aan de eis van maximaal 2 valse melding per dag. Om aan deze eis te voldoen kan onderzocht worden of het mogelijk is om de huidige eindfase van de detectiemethode te verfijnen. Ook kan er onderzocht worden of het gebruikte model signaal veranderd kan worden voor betere resultaten.

6.4 Aanbevelingen

Een van de eerste verbeteringen die uitgevoerd kan worden, is de calibratie voor de microfoons. In het huidige systeem kan er geen rekening gehouden worden met verschillende typen microfoons en met de sterkte waarmee ze ingesteld zijn. Dit is nodig, omdat de detectiemethode getraind wordt met een zorgvuldig samengesteld modelsignaal. En het modelsignaal wordt opgebouwd uit karakteristieke eigenschappen welke herleid zijn uit een audio opname. Als de microfoons van het systeem anders ingesteld zijn dan de microfoon die audio opname gemaakt heeft waaruit de karakteristieke eigenschappen herleid zijn, zal de detectie methode niet naar verwachting werken. Van de karakteristieke eigenschappen is de short term energy gevoelig voor de variatie van de instelsterkte van de microfoons. Indien de microfoon anders ingesteld wordt, dient het modelsignaal gecompenseerd te worden voor de verandering van het algemene

energieniveau.

Een andere optie die voor verbetering in de detectiepercentage kan bieden, is om de huidige resultaten verder te analyseren. Het huidige systeem detecteert puur op de uitkomst van het machine model, de gaussian mixture model, de support vector machine of de K-nearest neighbor. Er is nog geen rekening met hoe lang het detectie model een positief resultaat geeft gehouden en hoe vaak het resultaat afwisselt tussen positief en negatief. Kortom het gedrag van het huidige resultaat. Aan de hand van dit gedrag kan er een selectie methode ontwikkeld worden welke voor een hogere nauwkeurigheid zorgt. Er moet hierbij wel rekening gehouden worden met de maximale tijdsduur van 10 seconden van het rekenproces, het huidige systeem heeft een totale rekentijd van ongeveer 1 seconde.

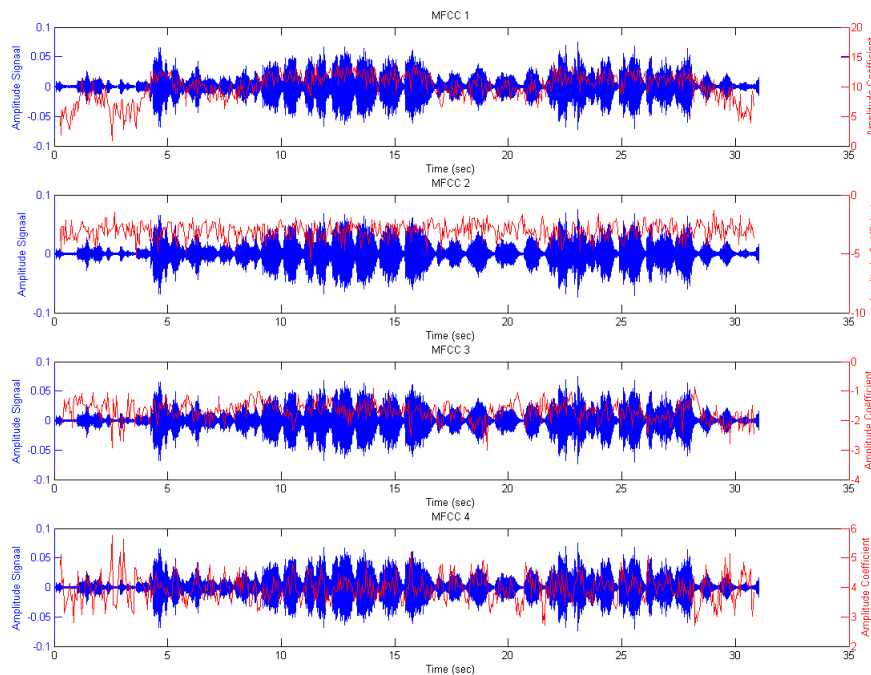
7 Literatuurlijst

1. G. Guo, S. Z. Li *Content-based audio classification and retrieval by support vector machines*, IEEE Transactions on Neural Networks, Vol. 14 (2003)
2. M. Maier, M. Hein, U. von Luxburg *Optimal construction of k-nearest-neighbor graphs for identifying noise clusters*, Theoretical Computer Science, Vol. 410, Issue 19, pp. 1749-1764. (2009)
3. E. Babaei, N. B. Anuar, A. W. A. Wahab, S. Shamshirband *An overview of audio event detection methods from feature extraction to classification*, Applied Artificial Intelligence, Vol. 31, issue 9-10, pp 661-714 (2017)
4. V. Exadaktylos, M. Silva, S. Ferrari, M. Guarino, C. J. Taylor, J. Aerts, D. Berckmans, *Time-series analysis for online recognition and localization of sick pig (Sus scrofa) cough sounds* Journal of the Acoustical Society of America Vol. 124, pp 3803-3809 (2009)
5. S. Chu, S. Narayanan, C.-C. J. Kuo *Environmental sound recognition with time-frequency audio features* IEEE Transactions on Audio, Speech and Language processing, Vol. 17, No 6 (2009)
6. J. Xie, M. Towsey, J. Zhang en P. Roe *Acoustic classification of Australian frogs based on enhanced features and machine learning algorithms*, Applied Acoustics. Vol 113, pp 192-201 (2016)
7. A. Temko, C. Nadeu *Classification of acoustic events using SVM-based clustering schemes*, Pattern Recognition, Vol 39. pp 682-694 (2006)
8. L. Lu, H-J. Zhang en H. Jiang *Content analysis for audio classification and segmentation*, IEEE transactions on Speech and Audio processing, Vol 10, No 7 (2002)
9. L. Lu, H-J. Zhang en H. Jiang *Content-based audio classification and segmentation by using support vector machines*, Multimedia Systems, Vol 8, pp 482-492 (2003)
10. S. Haykin en M. Moher *Introduction to analog & digital communications 2nd Edition*, Hoofdstukken 2,8 en 9, pp 18-91 313-391, John Wiley & Sons Inc, (2006)
11. M. D. Skowronski en J. G. Harris *Acoustic detection and classification of microchiroptera using machine learning Lessons learned from automatic speech recognition*, Journal of the Acoustical Society of America, Vol 119, pp 1817-1833 (2006)
12. S. Z. Li *Content-based audio classification and retrieval using the nearest feature line method*, IEEE transactions on Speech and Audio processing, Vol 8, pp 619-625 (2000)
13. J. P. Vink en F. de Haan *Comparison of machine learning techniques for target detection*, Artificial intelligence review, Vol 43, pp 125-139 (2012)
14. P. Dhanalakshmi, S. Palanivel, V. Ramalingam *Classification of audio signals using AANN and GMM*, Applied soft computing journal, Vol 11, pp 716-723 (2011)
15. T. Drugman *Using mutual information in supervised temporal event detection: Application to cough detection*, Biomedical signal processing and control, Vol 10, pp 50-57 (2014)

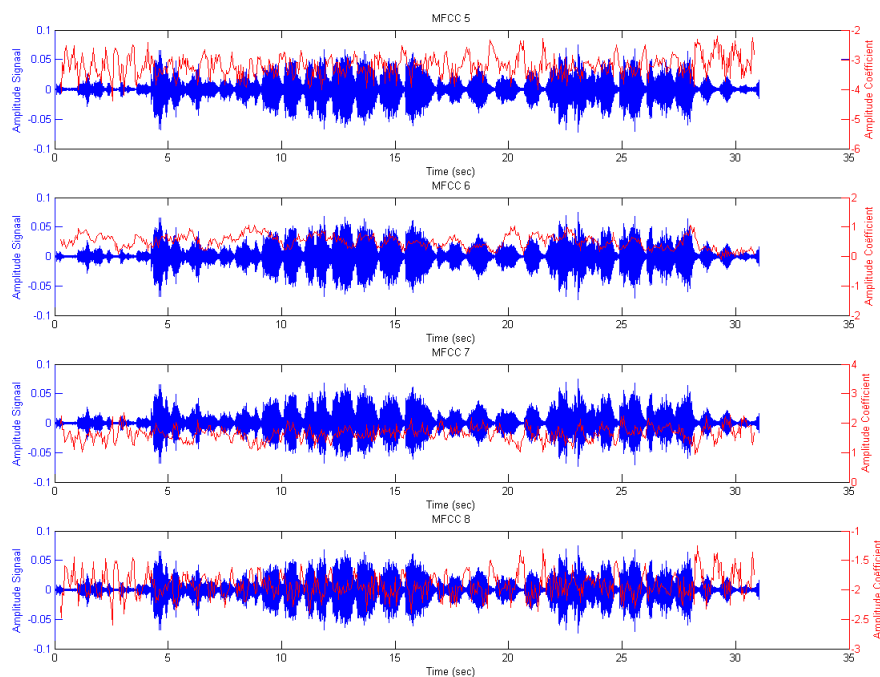
8 Bijlage

8.1 Bijlage 1: Karakteristieken van de Mel-frequency cepstrum coefficients (MFCC)

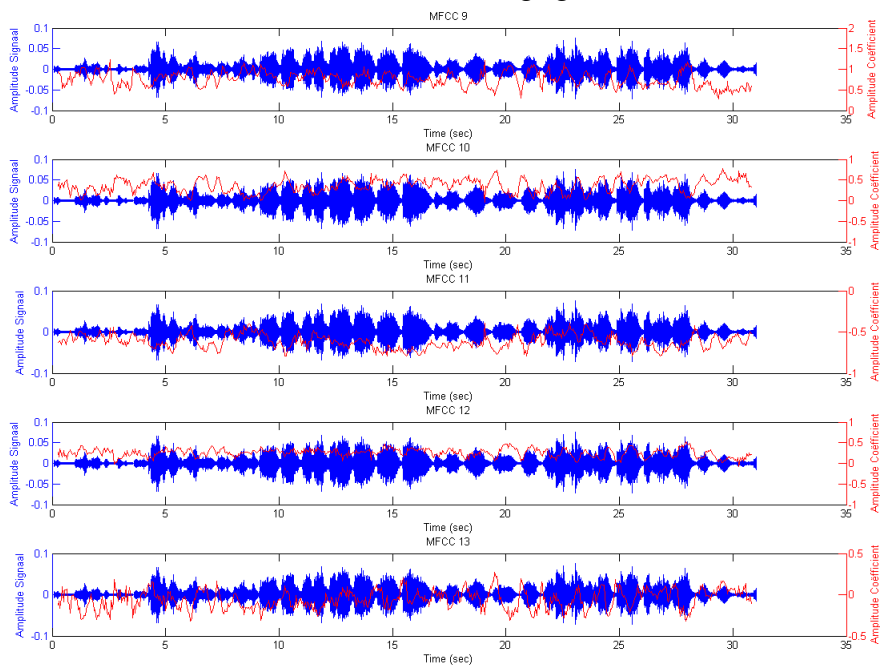
De MFCC karakteristieken van een doodskreet zijn weergegeven in figuur 29, figuur 30 en figuur 31. Ter vergelijking zijn de MFCC karakteristieken van ruis weergegeven in figuren 32, 33 en 34.



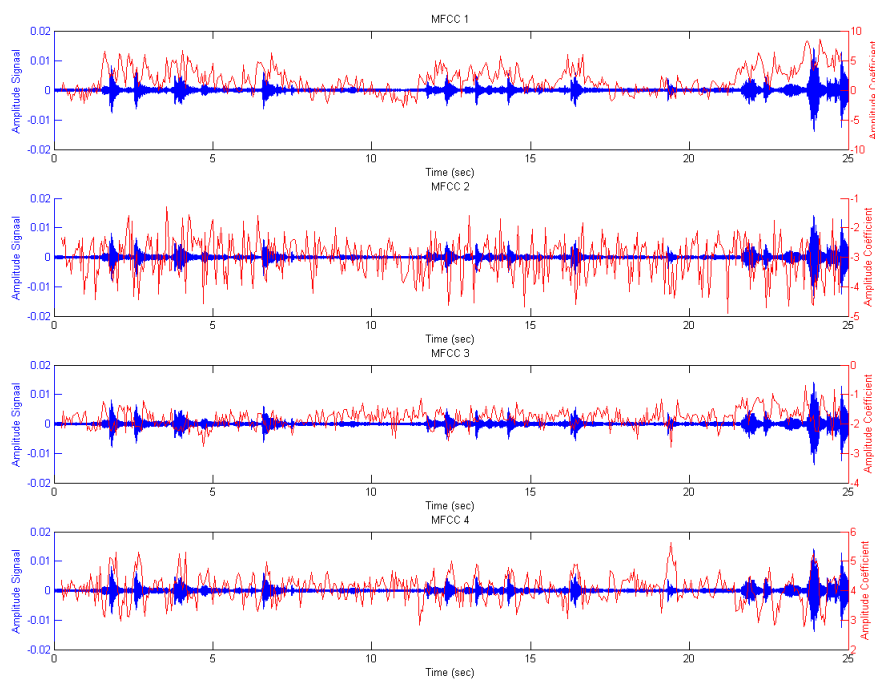
Figuur 29: Audio opname van een doodskreet, weergegeven in blauw, met de eerste 4 MFCC karakteristieken, weergegeven in rood



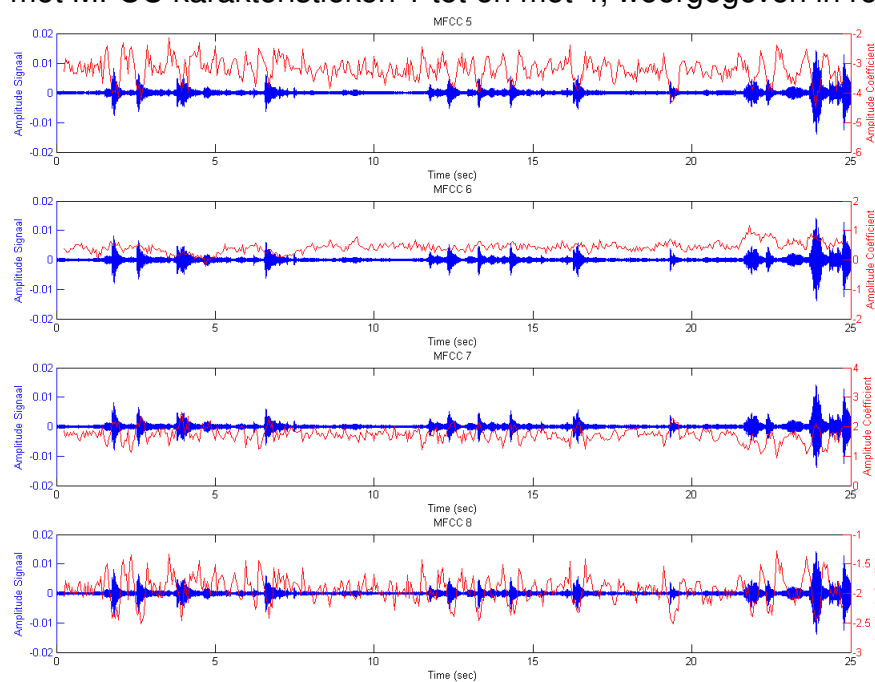
Figuur 30: Audio opname van een doodskreet, weergegeven in blauw, met MFCC karakteristieken 5 tot en met 8, weergegeven in rood



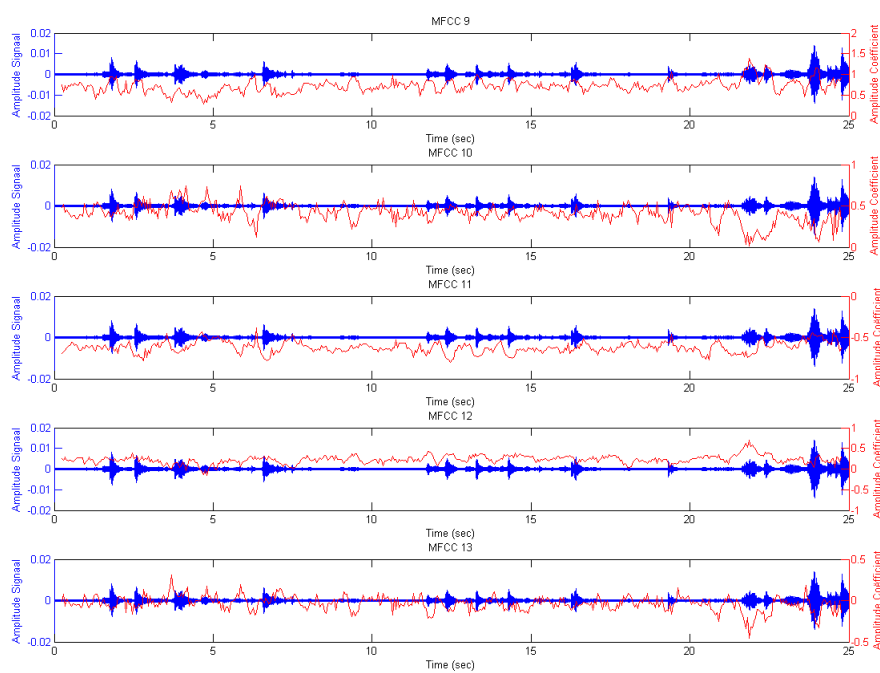
Figuur 31: Audio opname van een doodskreet, weergegeven in blauw, met MFCC karakteristieken 9 tot en met 13, weergegeven in rood



Figuur 32: Audio opname van achtergrondgeluid van een stal, weergegeven in blauw, met MFCC karakteristieken 1 tot en met 4, weergegeven in rood



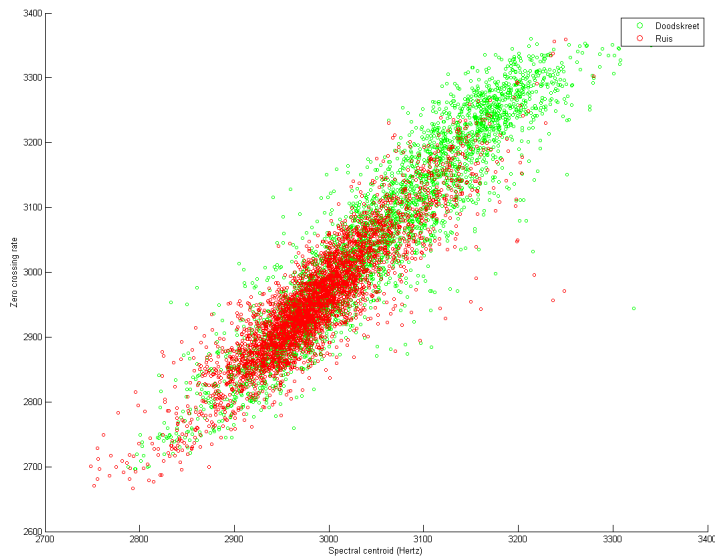
Figuur 33: Audio opname van achtergrondgeluid van een stal, weergegeven in blauw, met MFCC karakteristieken 5 tot en met 8, weergegeven in rood



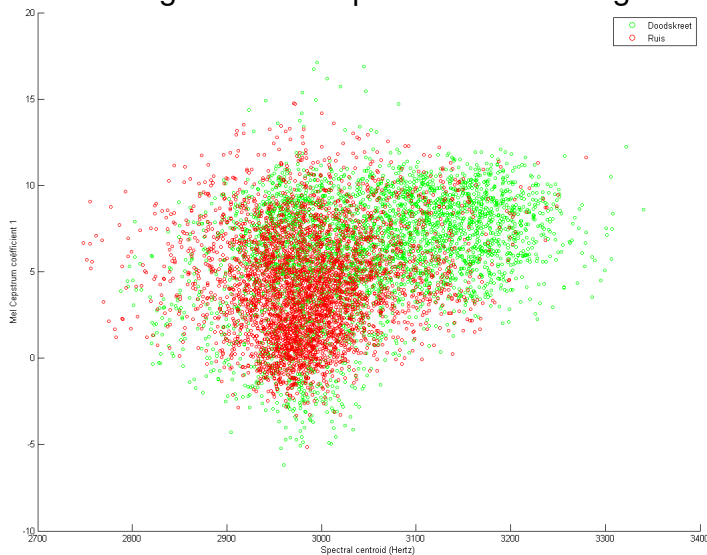
Figuur 34: Audio opname van achtergrondgeluid van een stal, weergegeven in blauw, met MFCC karakteristieken 9 tot en met 13, weergegeven in rood

8.2 Bijlage 2: Uiteenzetting van geluidsfeatures

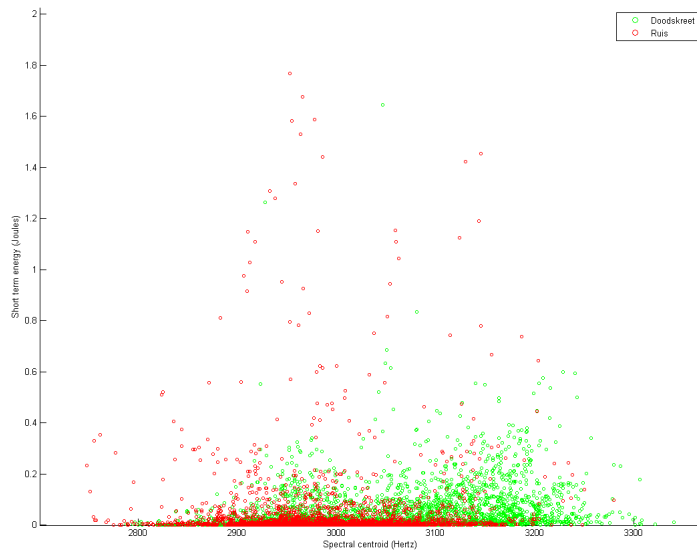
Verschillende geluidsfeatures zijn tegen elkaar uitgezet om de overlap tussen een doods-kreet en achtergrond geluiden te identificeren. De groene kleur is de locatie van de doodskreten en de rode kleur is locatie van de ruis signalen.



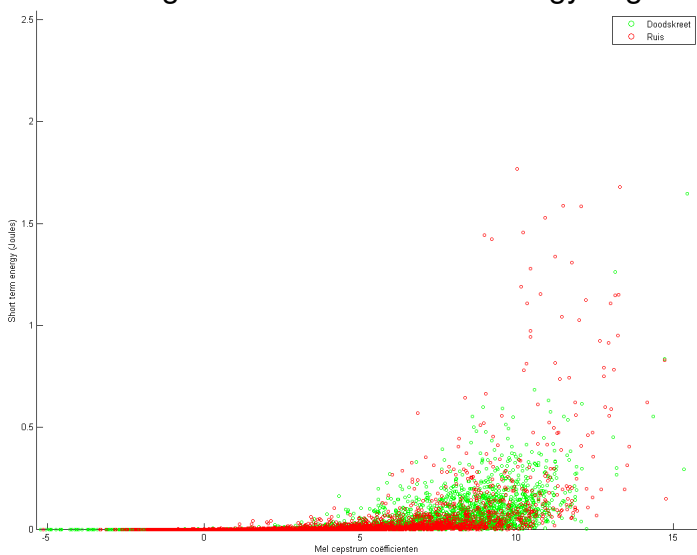
Figuur 35: De spectral centroid uitgezet tegen de zero crossing rate



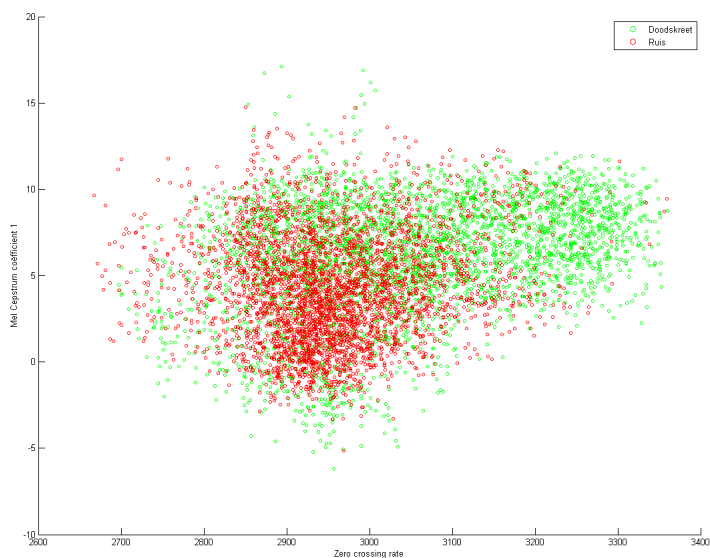
Figuur 36: De spectral centroid uitgezet tegen de MFCC 1 setting



Figuur 37: De shortterm energy uitgezet tegen de spectral centroid



Figuur 38: De shortterm energy uitgezet tegen de MFCC coefficienten



Figuur 39: de zero crossing rate uitgezet tegen de MFCC coefficienten