



# *Al en muziek*

*Prof. Dr. Ir. Bart De Moor*  
*ESAT-STADIUS, KU Leuven*

[www.alamirefoundation.org](http://www.alamirefoundation.org)  
[www.bartdemoor.be](http://www.bartdemoor.be)

# Overzicht

***I De wiskunde en fysica van muziek***

***II Muziek als taal***

***III Muziekherkenning en AI: Shazam***

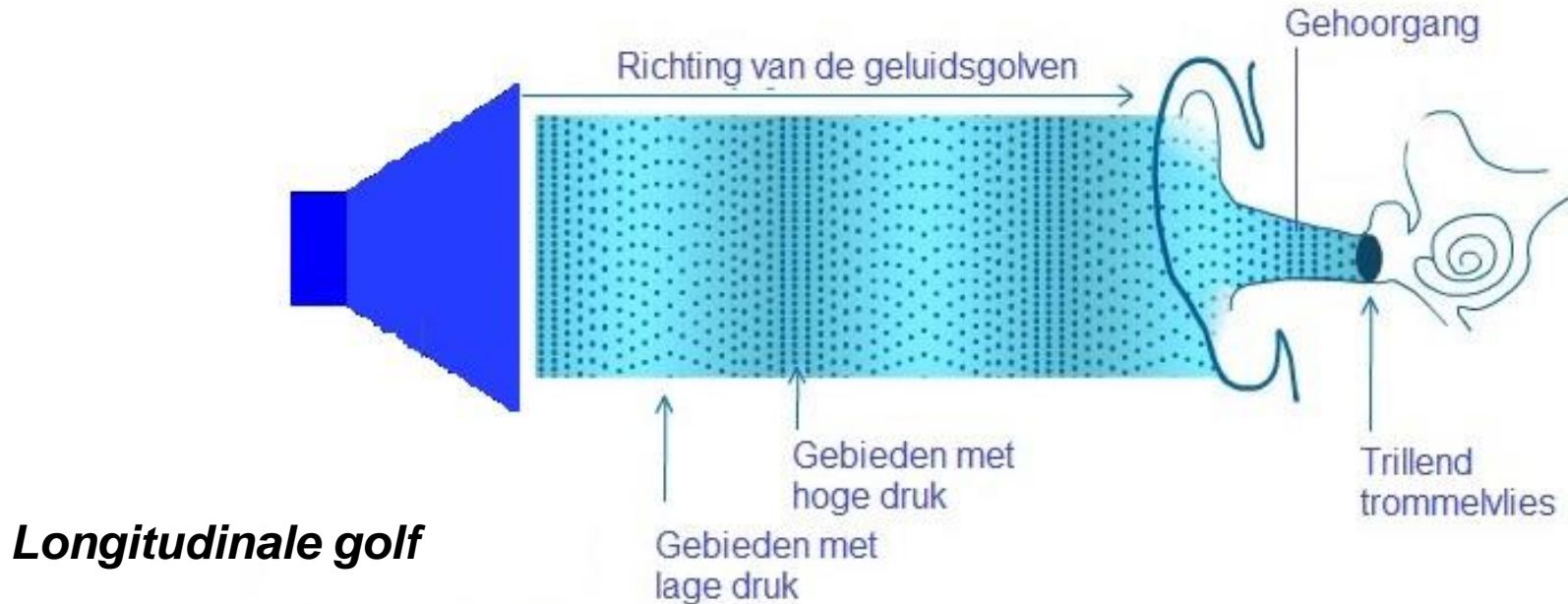
***IV AI en 'reverse acoustics'***

***V AI en polyfonie ontsluiten***

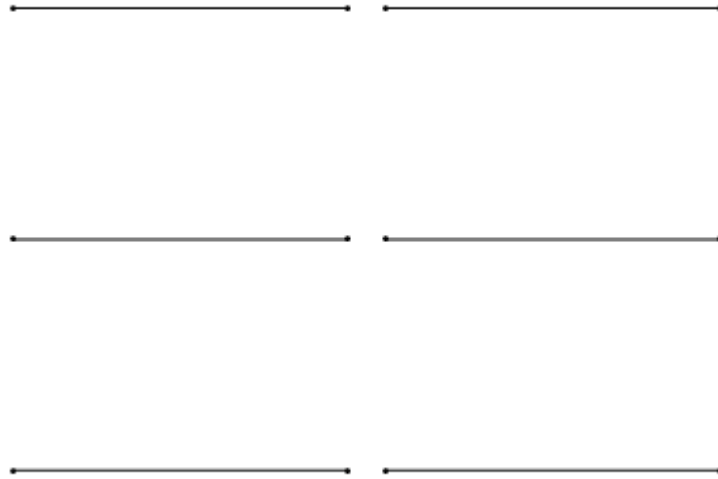
# ***I Wiskunde en fysica van de muziek***

# Wat is klank?

**Klank of Geluid = Trilling in een medium:**



# *De trillende snaar*

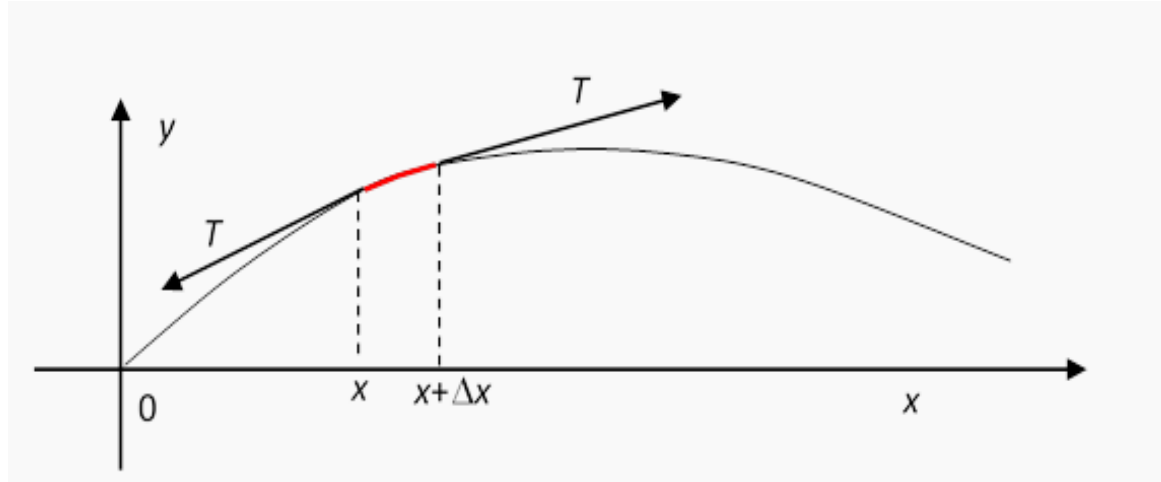


*Staande golven op een snaar*

*Fundamentele frequentie (toonhoogte/grondtoon) + eerste 5 overtonen (harmonischen)*

# De trillende snaar

*initiële uitwijking/  
aanslag*



**Spanning**

$$c^2 = \frac{T}{\rho}$$

**materiaaldichtheid**

**Bewegingsvergelijking  
voor een perfect elastische snaar :** 
$$\frac{\partial^2 y(x, t)}{\partial t^2} = c^2 \frac{\partial^2 y(x, t)}{\partial x^2}$$

# De trillende snaar

**resulterende trilling = fysische superpositie van oplossingen i.e. trillingen:**

$$y(x,t) = \sum_{n=1}^{\infty} \left[ (A_n \cos\left(\frac{n\pi ct}{L}\right) + B_n \sin\left(\frac{n\pi ct}{L}\right)) \right] \sin\left(\frac{n\pi x}{L}\right)$$



$$A_n = \frac{2}{L} \int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx. \quad y(x,0) = f(x)$$

$$B_n = \frac{2}{n\pi c} \int_0^L g(x) \sin\left(\frac{n\pi x}{L}\right) dx. \quad \frac{\partial y(x,t=0)}{\partial t} = g(x)$$

**$A_n$  en  $B_n$  functies van snaarlengte en initiële uitwijking.**





# De Trillende Snaar

**De oplossingsruimte aanpassen:**

$$y(x,t) = \sum_{n=1}^{\infty} \left[ (A_n \cos\left(\frac{n\pi ct}{L}\right) + B_n \sin\left(\frac{n\pi ct}{L}\right)) \sin\left(\frac{n\pi x}{L}\right) \right]$$

$$c^2 = \frac{T}{\rho}$$

**Wijzigen van snaarlengte  $L$**



**Wijzigen van spanning  $T$**



**Wijzigen van dichtheid  $\rho$**

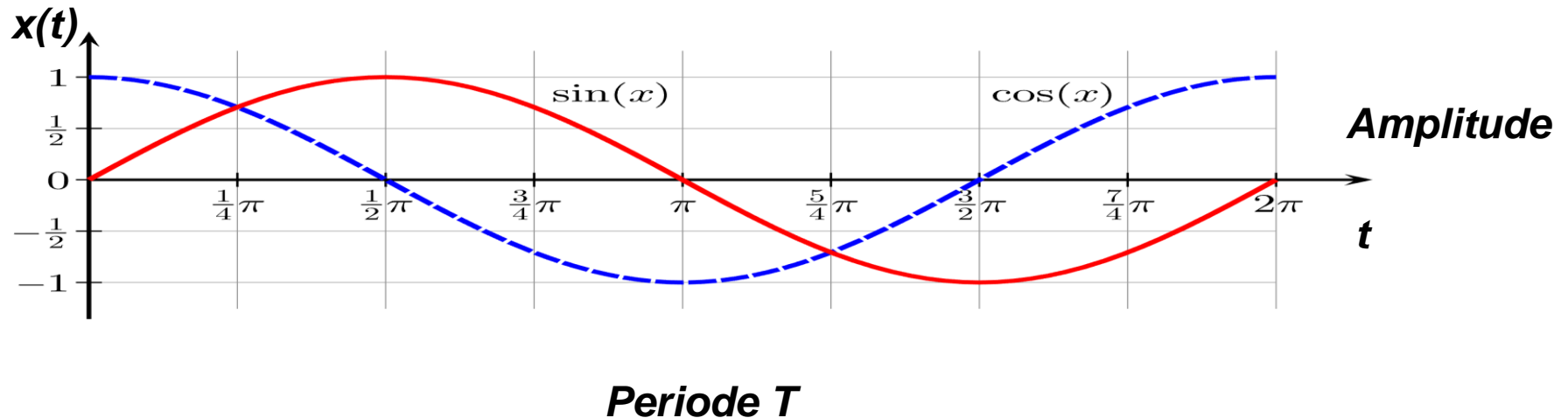


**runder-of schapendarm-snaar 9  
voor historische uitvoering**

# Stelling van Fourier

**Elke voldoende continue functie kan geschreven worden als som van sinussen en cosinussen.**

$$x(t) = a_0 + \sum_{k=1}^{\infty} (a_k \cos(2\pi kt/T) + b_k \sin(2\pi kt/T))$$



# *Stelling van Fourier*

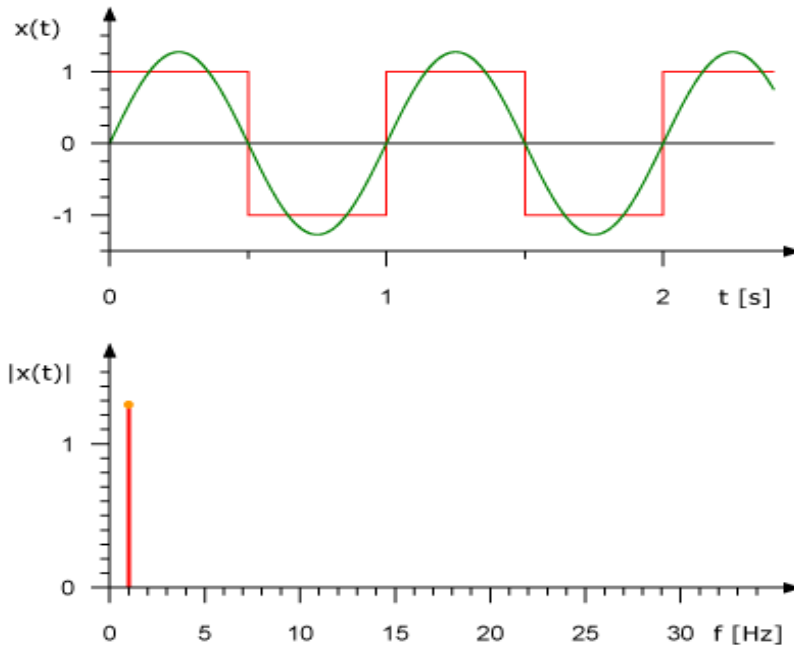
**Elke voldoende continue functie kan geschreven worden als som van sinussen en cosinussen.**

$$x(t) = a_0 + \sum_{k=1}^{\infty} (a_k \cos(2\pi kt/T) + b_k \sin(2\pi kt/T))$$

**Ideale snaar beschrijft het bijzondere geval waarbij de hogere harmonischen (golven voor verschillende  $k$ ) gehele veelvouden zijn van de fundamentele frequentie.**

# Fourierontbinding

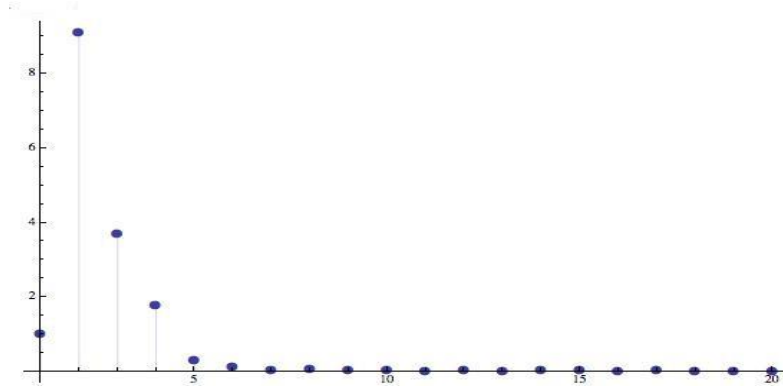
$$x(t) = a_0 + \sum_{k=1}^{\infty} (a_k \cos(2\pi kt/T) + b_k \sin(2\pi kt/T))$$



# Fourierontbinding

$$x(t) = a_0 + \sum_{k=1}^{\infty} (a_k \cos(2\pi kt/T) + b_k \sin(2\pi kt/T))$$

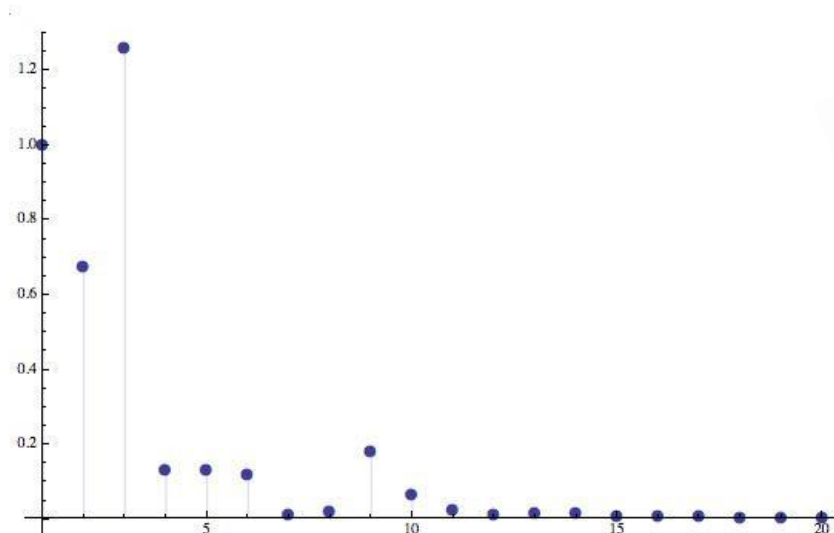
Amplitude



**zelfde La (A4)  
bij 440 Hz**

Harmonisch getal, k

Amplitude



**Verschillende instrumenten  
hechten bij een zelfde noot  
i.e. fundamentele frequentie  
een ander belang aan de  
overtonen.**

Harmonisch getal, k

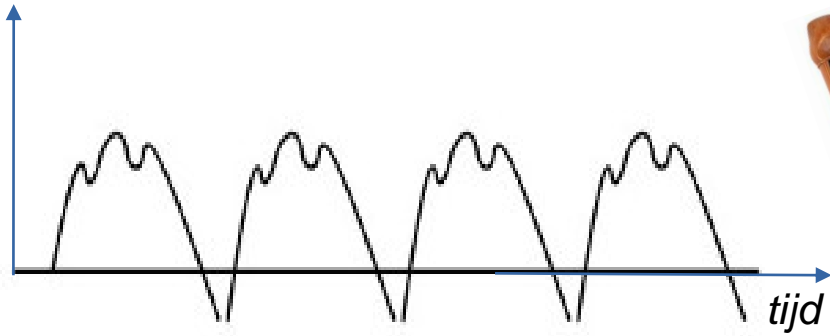
# Timbre/klankkleur

$$x(t) = a_0 + \sum_{k=1}^{\infty} (a_k \cos(2\pi kt/T) + b_k \sin(2\pi kt/T))$$

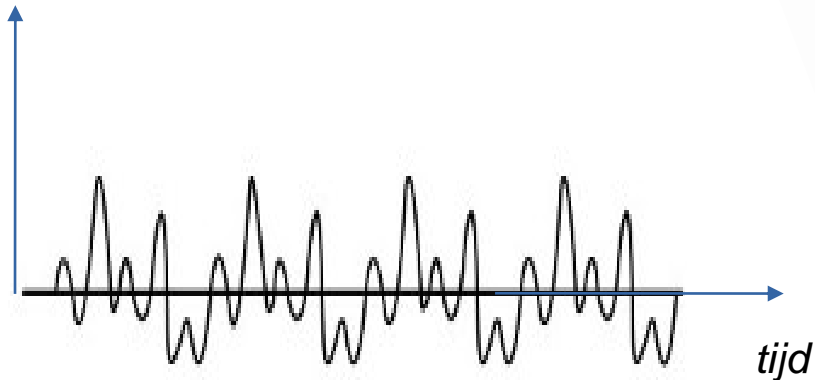
**zelfde La  
bij 440 Hz**

**Klankkleur is hoe we de vorm van het golfpakket i.e. de som van de golven in de Fourierontbinding interpreteren.**

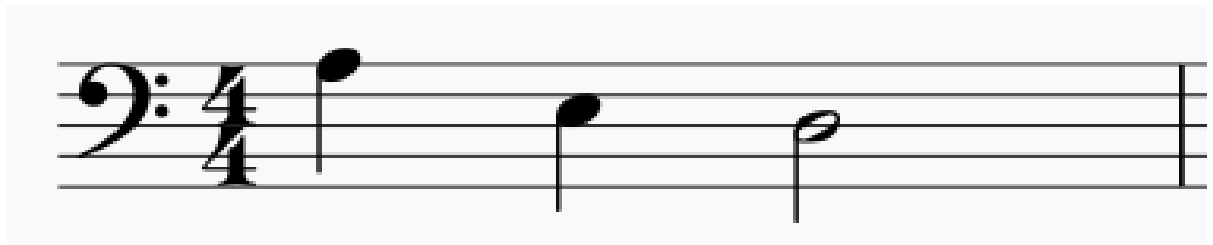
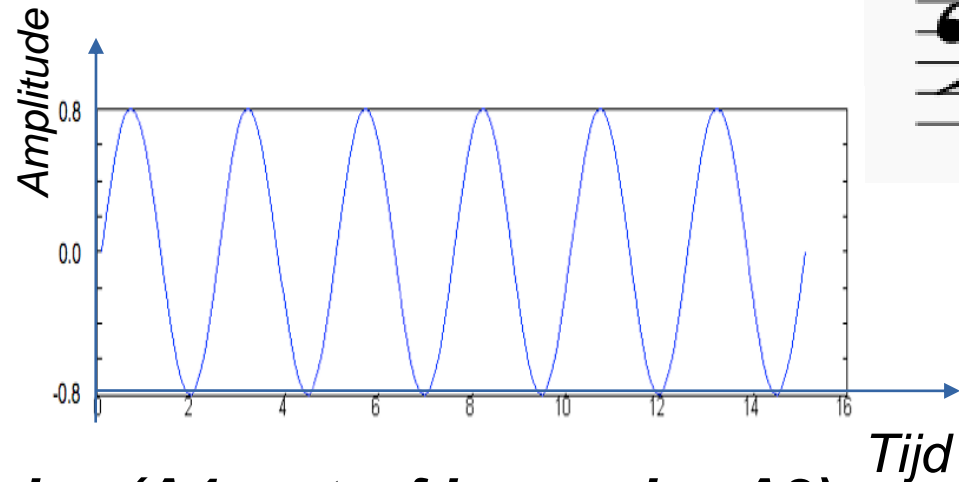
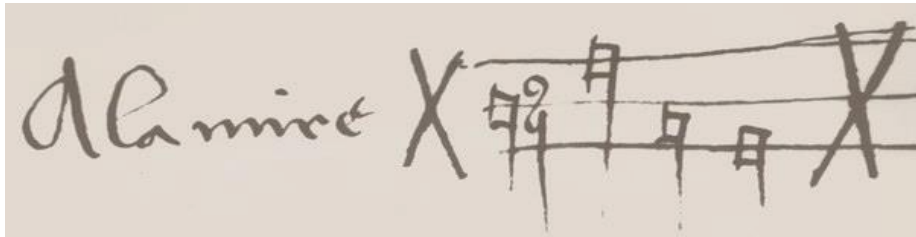
Amplitude



Amplitude



# Nootfrequenties



**La, A3**  
**220Hz**

**Mi**  
**164.81Hz**

**Re**  
**146.83Hz**

**La (A4, octaaf hoger dan A3)**  
**440HZ = 440 pieken per seconde**

# Nootfrequenties: verhoudingen

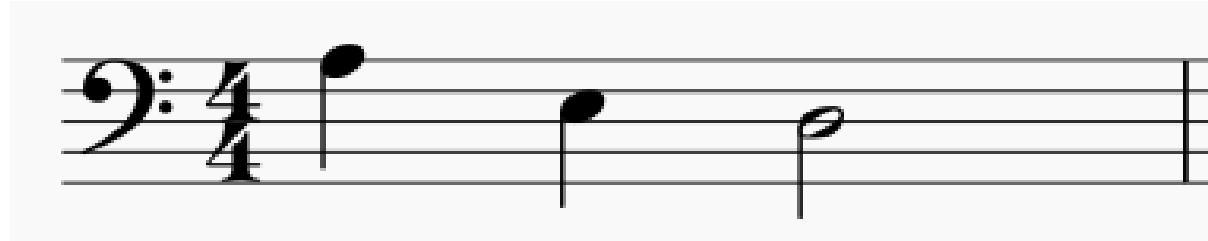
## Frequentieverhouding grondtonen

$A4/A3 \approx 2/1$  octaaf

$A4/E3 \approx 4/3$  reine kwart

$A4/D3 \approx 3/2$  reine kwint

$E3/D3 \approx 9/8$  grote secunde



La, A3  
220Hz

Mi, E3  
164.81Hz

Re, D3  
146.83Hz

**Consonante intervallen zijn die intervallen met een eenvoudige verhouding tussen de frequentie van de grondtonen.**



# ***II Muziek als taal***

# ***Wat is muziek?***

*Muziek, van het grieks μουσική (mousikè), de kunst van de muzen.*

*De kunstzinnige schikking en combinatie van de klanken van muziekinstrumenten en de menselijke stem om schoonheid van vorm dan wel uitdrukking van emotie te bereiken (Wikipedia).*



# Globale muzikale structuur

Ritme, Toonaard, Tempo,  
atmosfeer

*Des pas sur la  
neige*  
- Debussy

Triste et lent (♩ = 44)

*pp* *p expressif et douloureux*

*più pp*

*Ce rythme doit avoir la valeur sonore  
d'un fond de paysage triste et glacé*

Gestructureerde onderverdeling

bv. Sonata  
vorm:

Exposition Development Recapitulation

A: I B: V V (and others) A: I B: V

minor: III or V

# Lokale muzikale structuur

## Melodische structuur

**Monofonie** =  
een enkele  
muzikale lijn in  
unisono

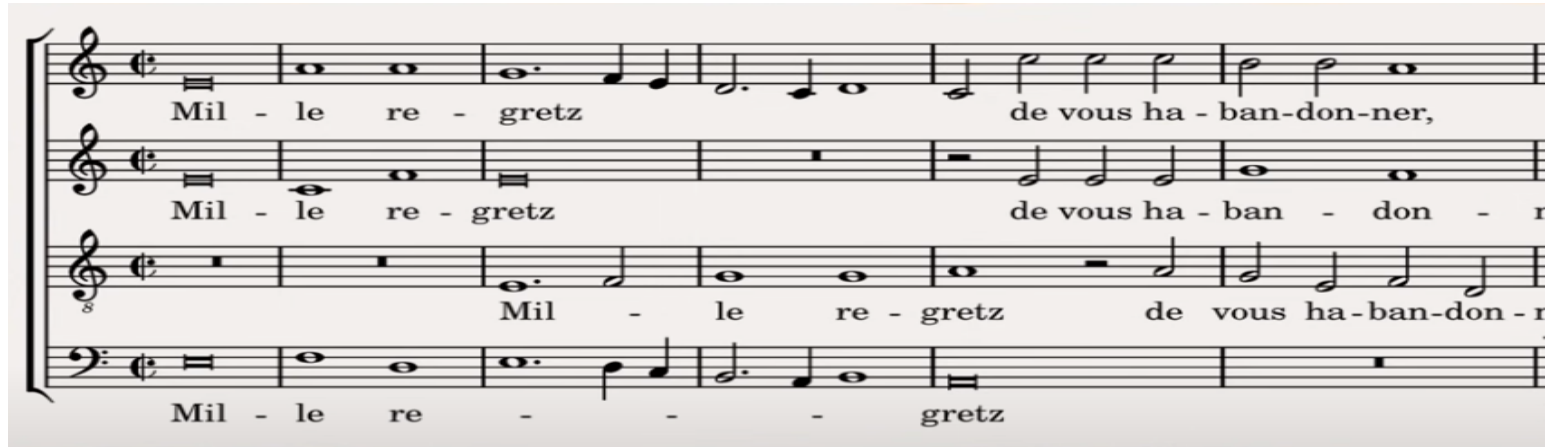
Resp.  
8.  
I



n pa- ce in i- dipsum \* Dórmi- am et re-

Het responsorium *In pace in idipsum*

**Polyfonie** =  
verschillende  
melodieën in  
contrapunt



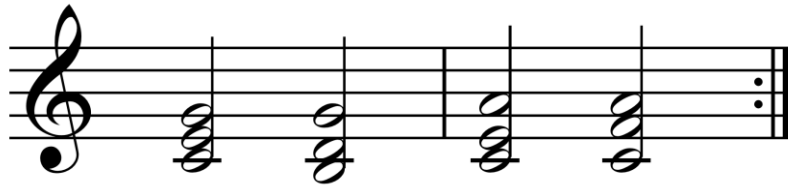
Mil - le re - gretz de vous ha - ban - don - ner,  
Mil - le re - gretz de vous ha - ban - don - ner,  
Mil - le re - gretz de vous ha - ban - don - ner,  
Mil - le re - gretz

*Mille regretz, Josquin des Prez*

# Lokale muzikale structuur

## Akkoorden en Harmonische progressies

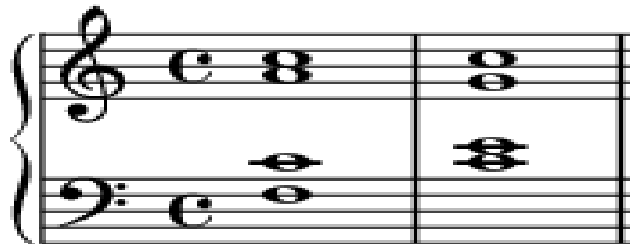
I-V-vi-IV



*Let it be, Can you feel the love tonight, Take on me, Don't stop believin', ...*

## Cadenzen

IV-I  
Amen cadens



# ***III Muziekherkenning en AI: Shazam***

# *Shazam*

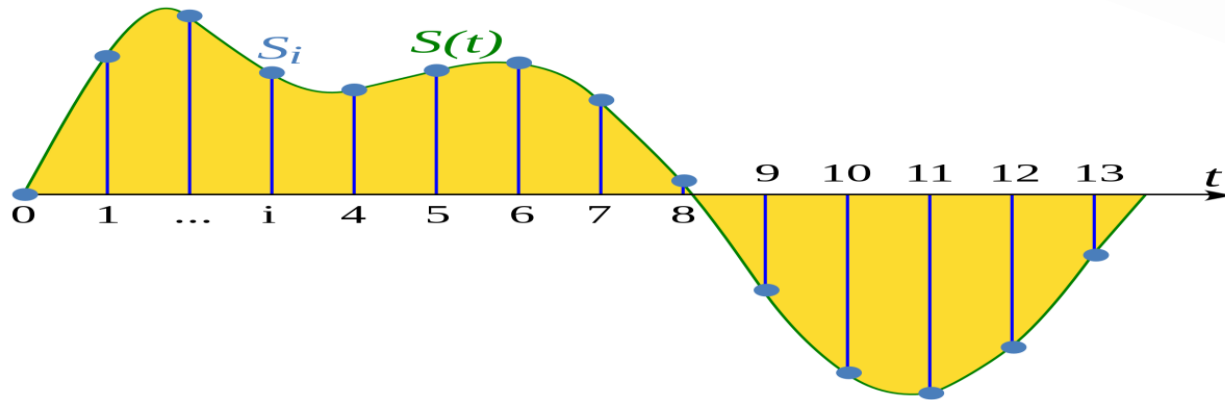


- Audioherkenning op basis van korte fragmenten (5 s) ook in rumoerige omgeving
- Database van 8 miljoen audiobestanden, van muziek tot reclamedeuntje
- Meer dan 1 miljard app downloads
- Hoeveelheid queries is graadmeter voor populariteit/inkomsten artiesten

# Hoe werkt Shazam?



**Van digitaal naar analoog:**



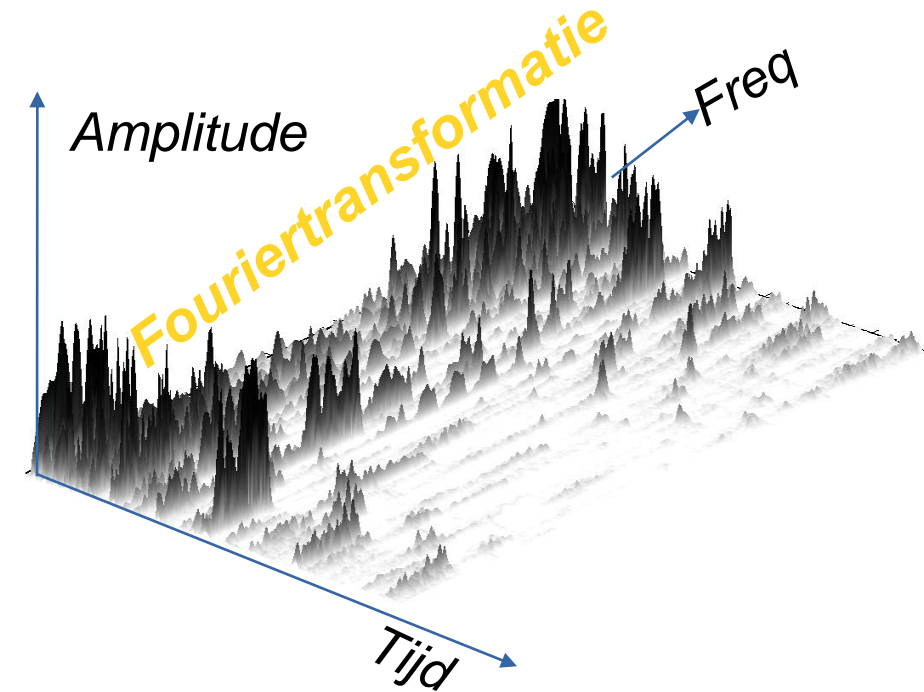
**Nyquist-Shannon:**

*Voor foutloze bemonsteringsfrequentie  $\geq 2 \times$  maximale frequentie in het signaal  
=> Bemonsteringsfrequentie CD, MP3, ... op  $2 \times$  maximaal hoorbare freq  $\sim 20$  kHz*

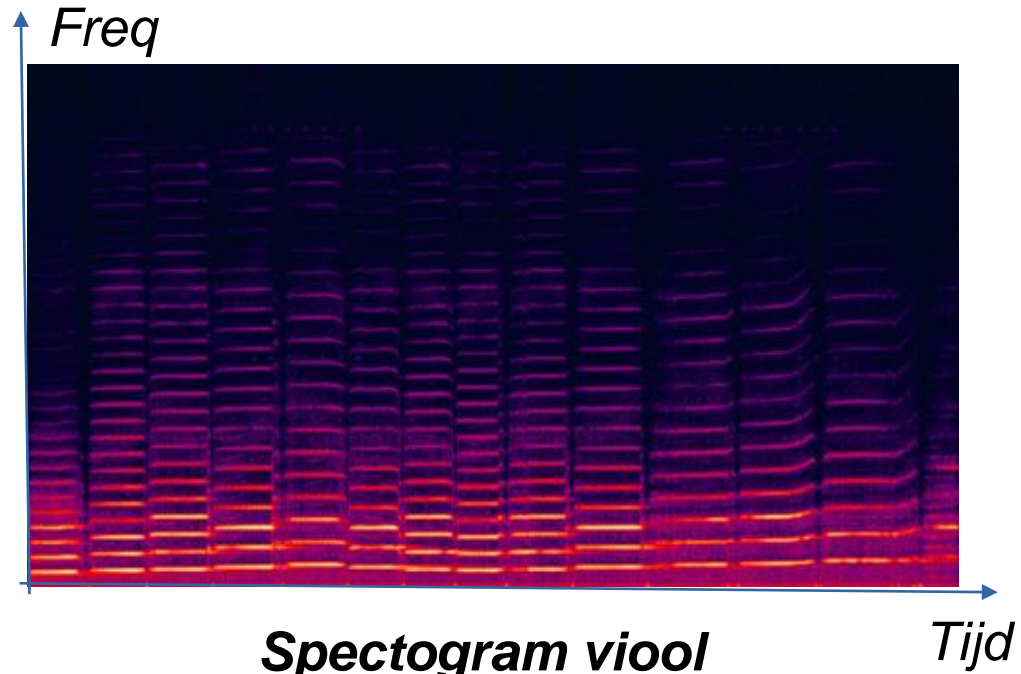


# Hoe werkt Shazam?

**Spectrogram: een muzikale vingerafdruk**



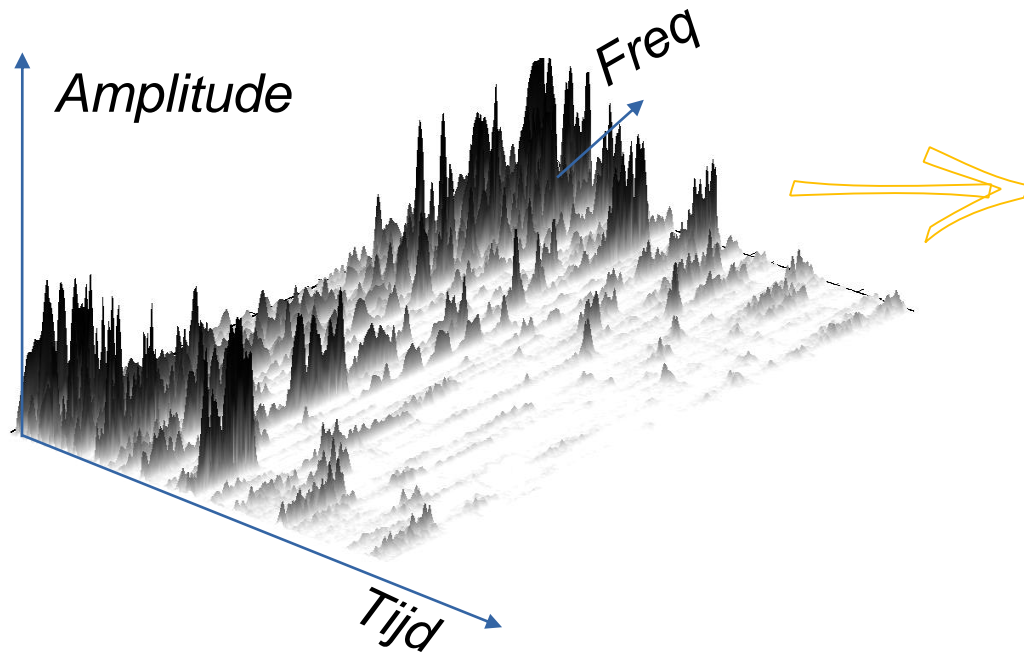
**Spectrogram muziekstuk**



**Spectrogram viool**  
*Piekintensiteit bij fundamentele frequenties!*<sub>25</sub>

# Hoe werkt Shazam?

**Spectrogram: een muzikale vingerafdruk**



**Spectrogram muziekstuk**

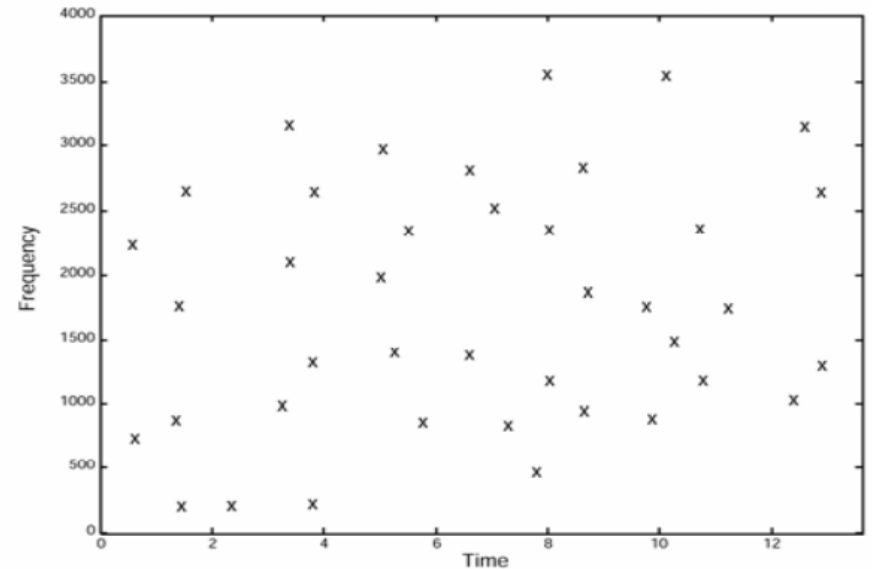


Fig. 1B - Constellation Map

**Constellatiekaart** : Ruis filteren door uit het spectrogram piekamplituden te kiezen

# Hoe werkt Shazam?

## Constellatiekaart

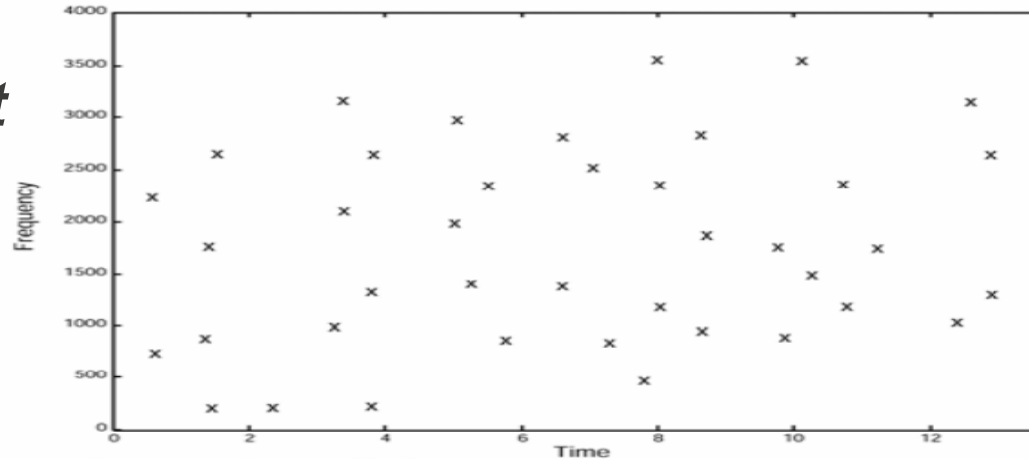


Fig. 1B - Constellation Map

- Wordt **bij query gemaakt** over lengte opname van de gebruiker
- **Werd gemaakt** over gehele tijdsduur van alle stukken in Shazam **databank**

***Song matching***

***//***

***constellatiekaart van query komt overeen met een kaart uit databank***

# Hoe werkt Shazam?

**Probleem:** één song bevat ettelijke constellatiekaarten en enkele piek is weinig informatief → **vergelijking constellatiekaart query en databasekaarten duurt lang.**

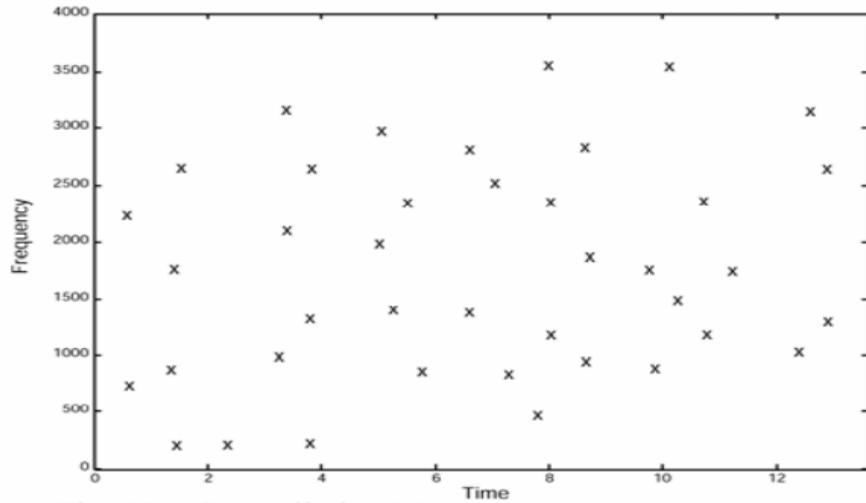


Fig. 1B - Constellation Map

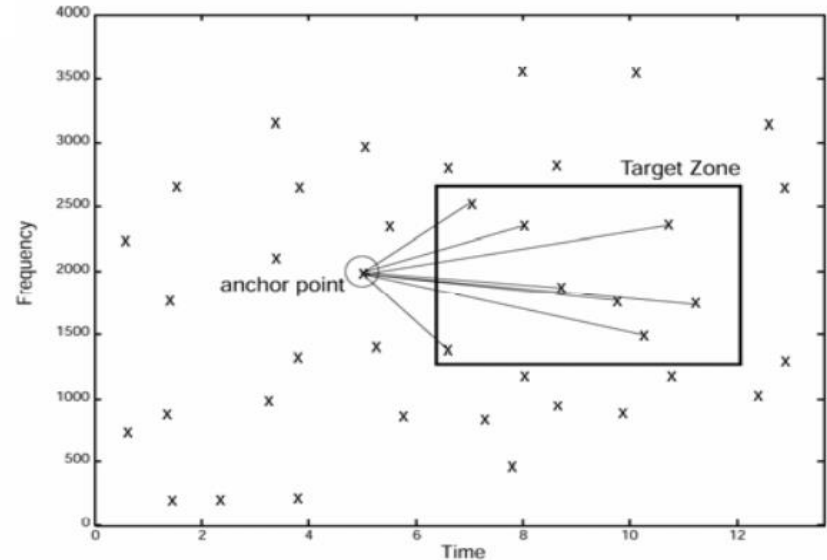


Fig. 1C - Combinatorial Hash Generation

**Oplossing:** sample in regelmatige tijdsintervallen een ankerpunt (piek) en verbind deze met andere pieken in naburige regio in de constellatiemap.

→ **minder pieken om te vergelijken, wél meer informatief**

# Hoe werkt Shazam?

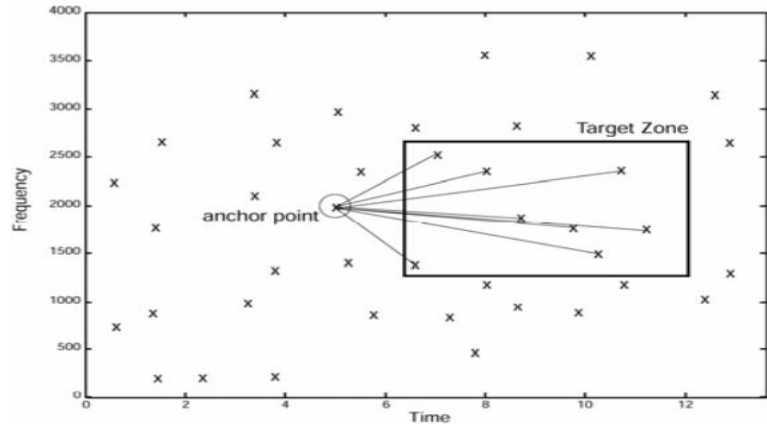


Fig. 1C - Combinatorial Hash Generation

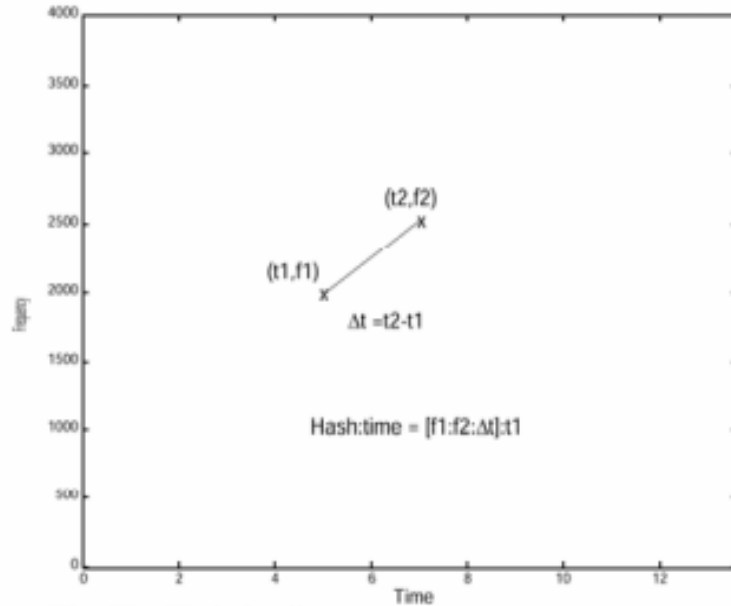
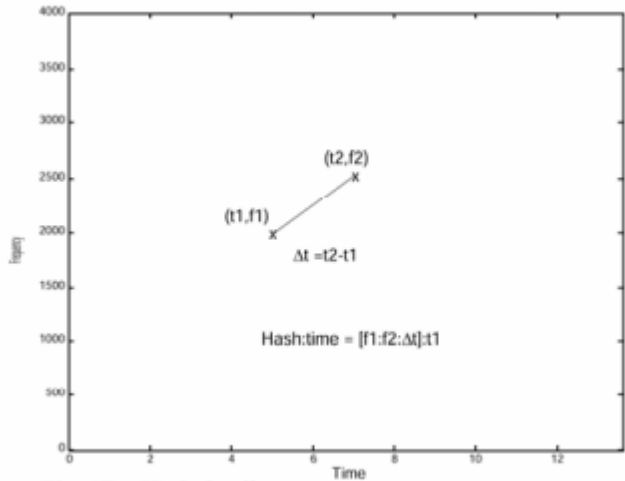


Fig. 1D - Hash details

Een koppel punten = **hash** uit het constellatiediagram van de query bestaat uit [hoogte piek 1, hoogte piek 2, tijdsverschil tussen pieken]

# Hoe werkt Shazam?

*bitrepresentatie hash*

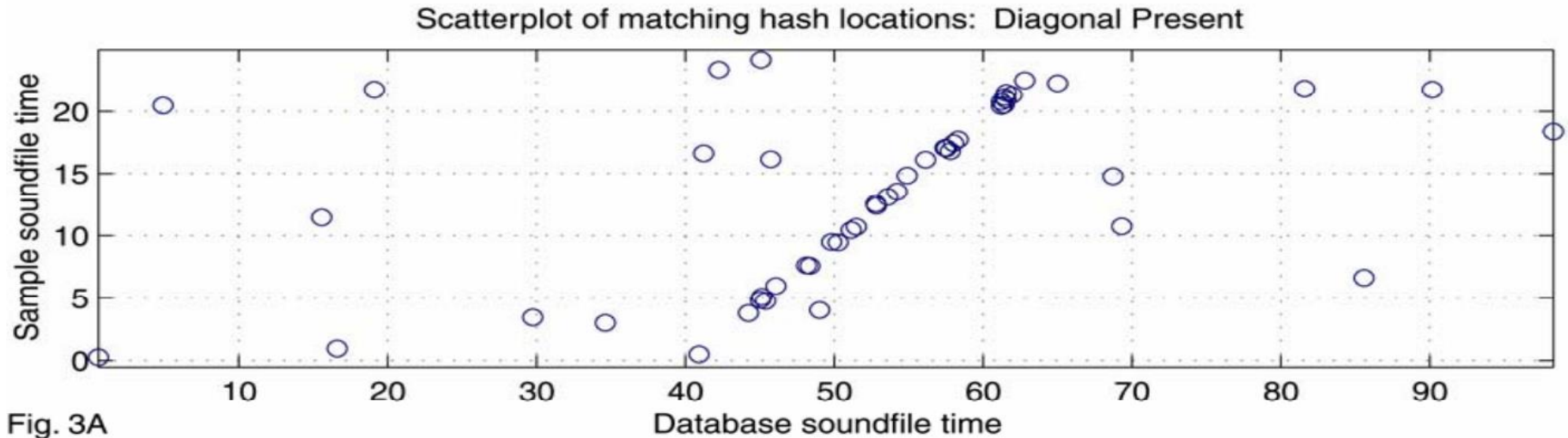


Hash Tag	Time in Seconds	Song
30 51 99 121 195	53.52	Song A by artist A
33 56 92 151 185	12.32	Song B by artist B
39 26 89 141 251	15.34	Song C by artist C
32 67 100 128 270	78.43	Song D by artist D
30 51 99 121 195	10.89	Song E by artist E
34 57 95 111 200	54.52	Song A by artist A
34 41 93 161 202	11.89	Song E by artist E

Ook een hash uit het constellatiediagram van de database bestaat uit: [hoogte piek 1, hoogte piek2, tijdsverschil tussen pieken ]

Met deze hash associëren we de aanvangstijd van het ankerpunt en het het werk waaruit de constellatiemap komt.

# Hoe werkt Shazam?



Opname kan op eender welk moment in het afgespeelde nummer plaatsvinden

Voor hashes is echter de lengte van de tijdsintervallen van belang, niet de absolute tijd

→ ***Match als hash van sample en hash database file op een diagonaal liggen***

# ***IV AI en 'reverse acoustics'***



# *Probleem: capteren en reproducieren van akoestiek*



***Meting***



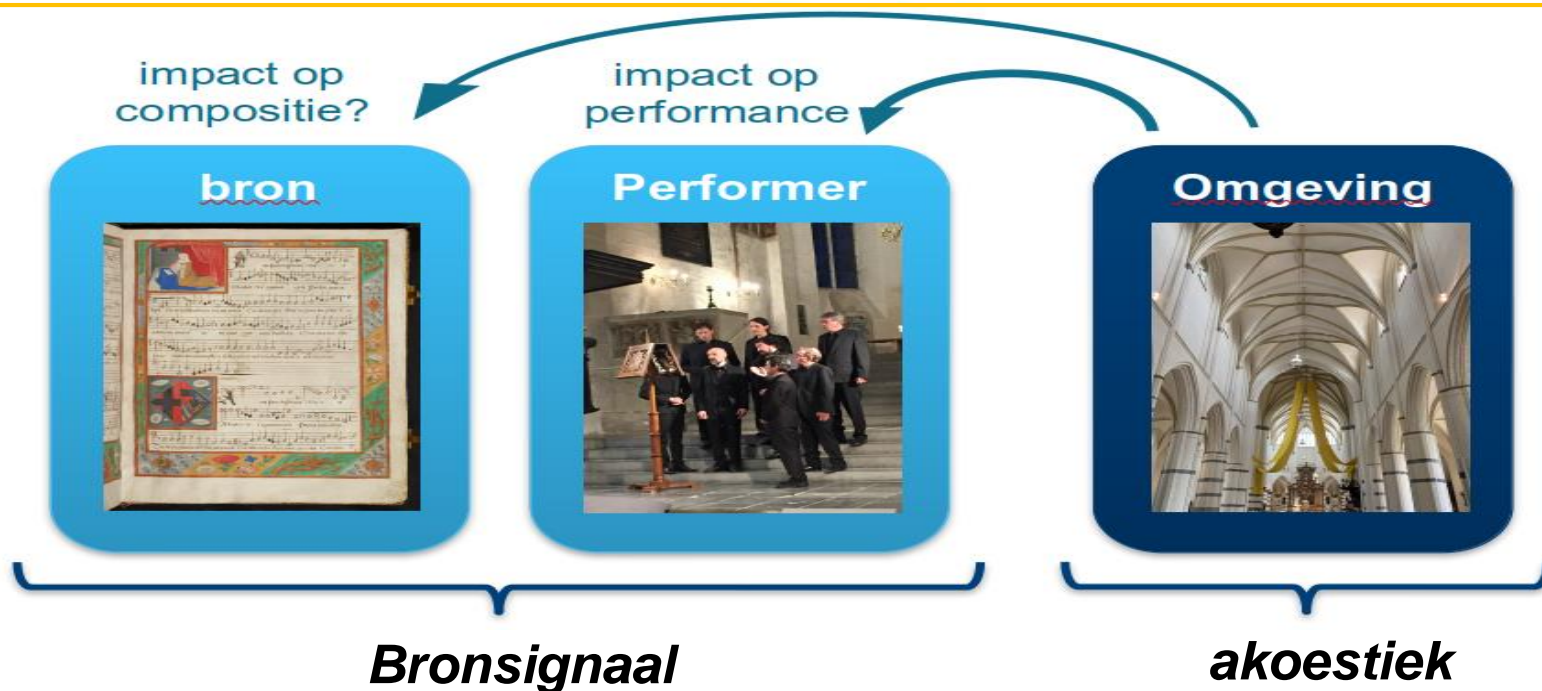
***Archivering***



***Reproductie***

# Wat is de akoestiek van een ruimte?

**Akoestiek = voortplanting en reflectie van geluidsgolven binnen een gegeven ruimte onafhankelijk van het bronsignaal**

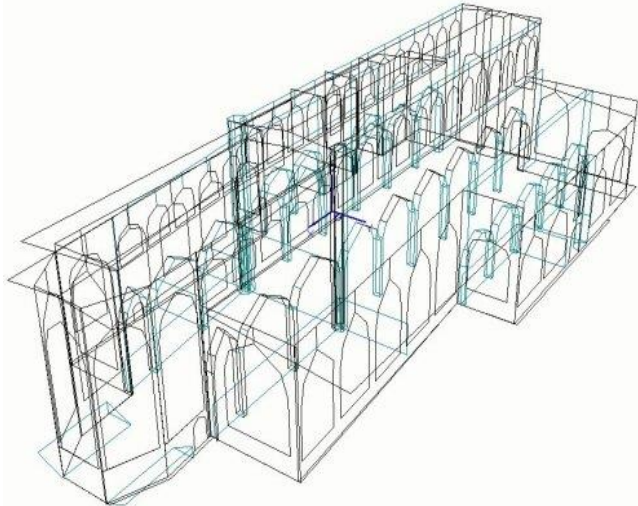


# Ruimtelijke akoestiek capteren



1. Meet geometrie van ruimte en objecten

2. Meet verstrooiingseigenschappen van ruimte en objecten



→ *Reproductie?*

→ Moeilijk, tijdsintensief, duur

# Data-driven modeling

## Fundamenteel resultaat voor lineaire tijds-invariante systemen



delta functie

impuls respons

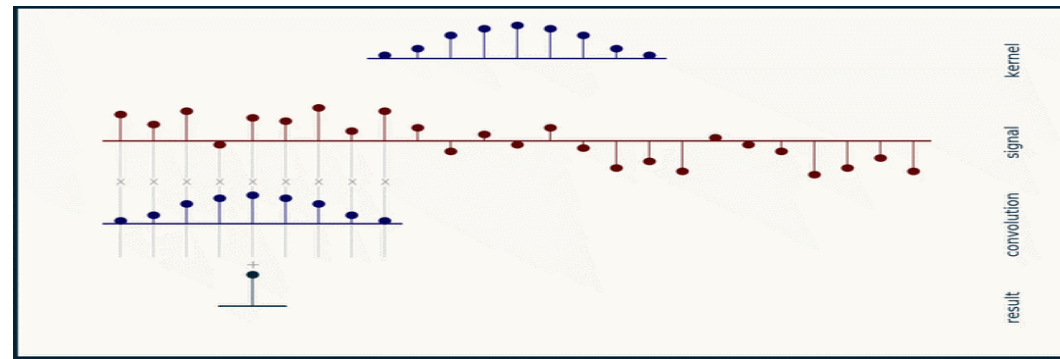
*Impulsrespons bevat alle informatie van het systeem.*

*Output systeem met willekeurige input kan berekend worden via convolutie*

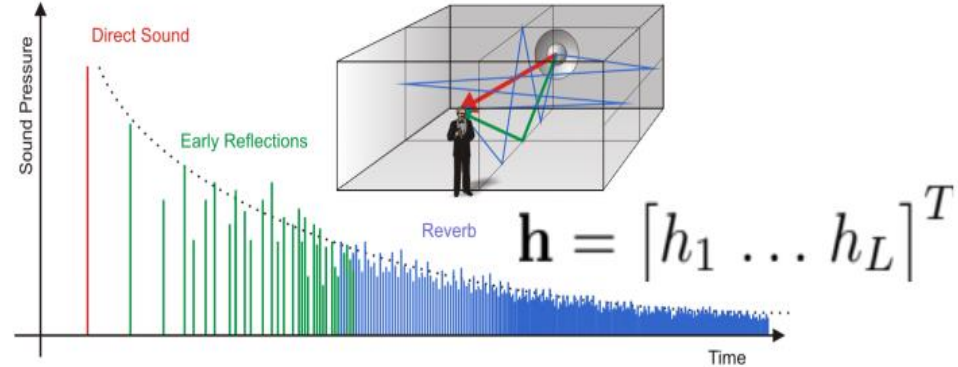
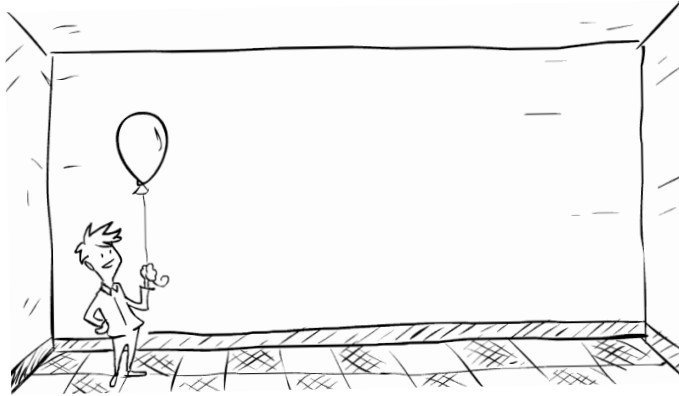
$$\mathbf{h} = [h_1 \dots h_L]^T$$

$$\mathbf{x} = [x_1 \dots x_N]^T$$

$$\mathbf{y} = \mathbf{h} * \mathbf{x}$$



# Data-driven modeling



Pieken in impulsrespons representeren akoestische reflecties

Data-driven model voor ruimteakoestiek

**Probleem:** 1 impulsrespons  $\sim 10^4$  samples  $\sim 10^5$  bits

Plenkoestische sampling theorie:

$100^3$  bron positie  $\times 100^3$  obs positie

Grootte orde = petabyte (1 petabyte = +/- 200 000 movie DVDs)

# Meting: data-driven modeling

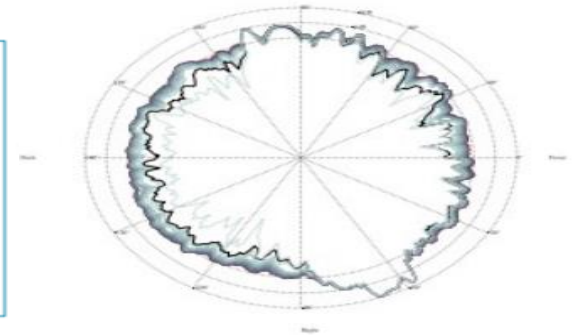
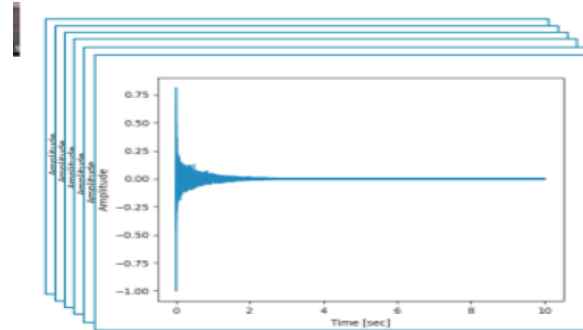
## Case study: Nassau Chapel, KBR Brussel



Genelec 8030C



G.R.A.S. VIP 50VI-1



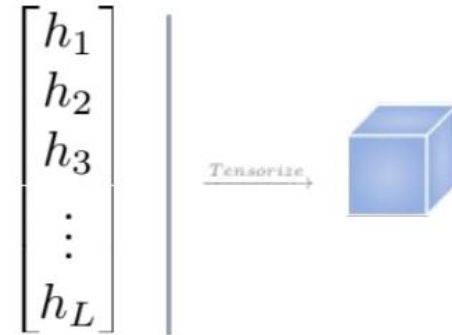


# AI en archivering: data-approximatie

Impulsrespons = vector in  $10^4$  dimensies  $\rightarrow$  **Curse of Dimensionality**

Hervorm de vector ( 1D rij ) naar N-D rij:

- *Matrix* = 2 - D rij  $\sqrt{L} \times \sqrt{L} = 100 \times 100$
- *Tensor* = N- D rij  $\sqrt[N]{L} \times \sqrt[N]{L} \times \dots \times \sqrt[N]{L}$



Tensor valt te ontbinden en te benaderen in componenten (polyadische decompositie) :



# Reproductie



→ 3D luidsprekeropstelling in semi-anechoïsche kamer

## Alamire Interactive Lab @ Library of Voices

→ Convolutie van het bronsignaal met de ruimtelijke impulsrespons waar reflectiehoek  $\approx$  luidsprekerhoek

→ **REAL-TIME** emuleren van ruimtelijke akoestiek



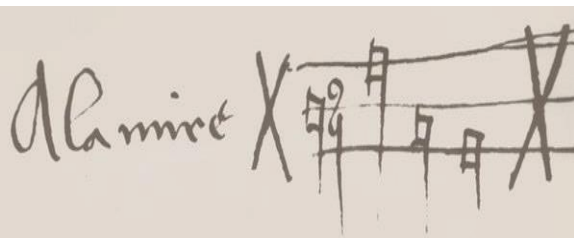


# ***V AI en polyfonie ontsluiten***

# Ontsluiten van polyfonie



- Franco-Flamands gedurende de renaissance bij de **meest gewaardeerde musicici en componisten** in de Westerse wereld.
- Hun muziek, de **polyfonie**, is genoteerd in mensurale notatie, een in onbruik geraakt systeem, door velen **moeilijk te interpreteren** → **verminderde verspreiding**
- Veel van dit erfgoed dient nog ontsloten te worden door **digitizatie** van tekst, muziek, illuminaties en heraldiek en het maken van **transcripties**.



# Alamire Foundation



Internationaal studiecentrum in samenwerking met KU Leuven, voor muziek in de Nederlanden van de middeleeuwen tot 1800

Stimuleert en voert wetenschappelijk- en praktijkonderzoek uit  
Bijzondere aandacht voor onbekende of onontgonnen bronnen

Ontwikkelt state-of-the-art methodes voor digitalizatie, valorisatie en conservatie van muzikaal erfgoed



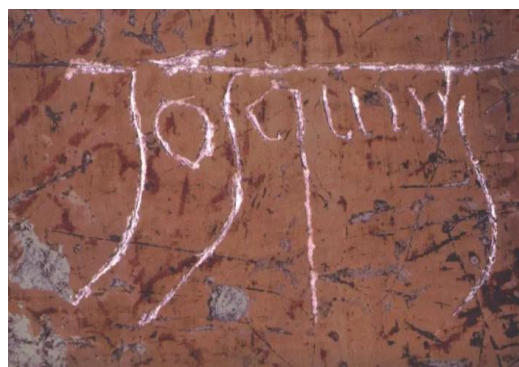


# Alamire Foundation



Verspreiding, conservatie en studie van Vlaams erfgoed

Franco-Flamands zoals Dufay, Ockeghem, de Rore en Josquin Des Prez als muzikale Vlaamse Primitieven: absolute wereldtop!



# Ontsluiten van polyfonie

## Mensurale notatie

Notatiesysteem voor meerstemmige muziek in de renaissance

Geen maatstrepen, enkel *dubbele maatstreep* op het einde van de partij

Stemmen zijn niet gealigneerd, dus geen onmiddellijk overzicht op harmonische structuur

Mollen en kruisen niet altijd genoteerd: *musica ficta*



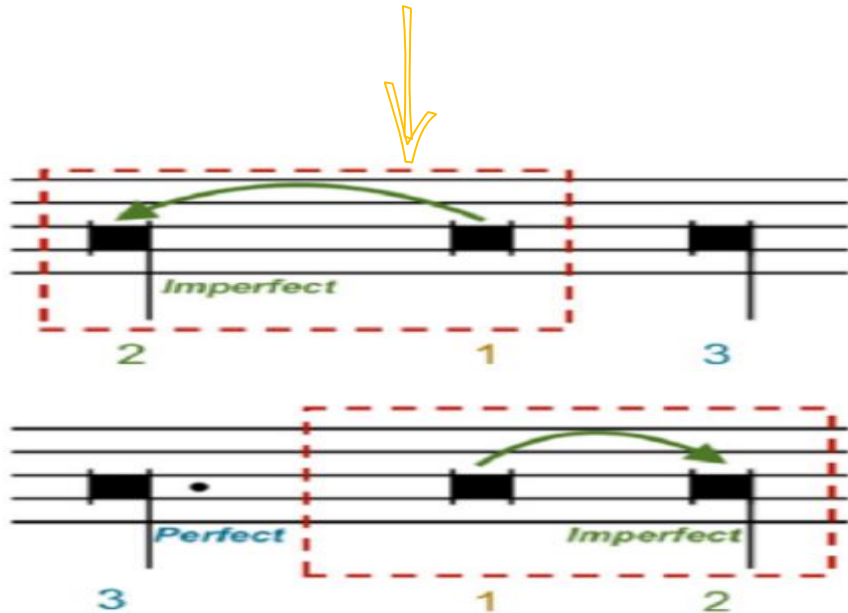
**Leuven**  
**Chansonnier**



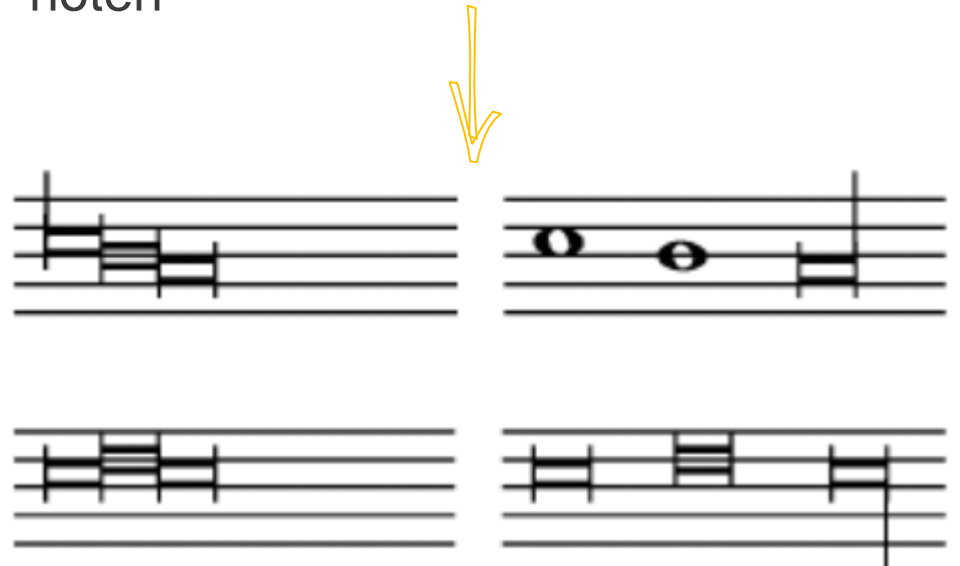
# Ontsluiten van polyfonie

## Ambigüiteit mensurale notatie

punt kan verdelen in groepen van drie  
punt kan ook duratie met helft verlengen

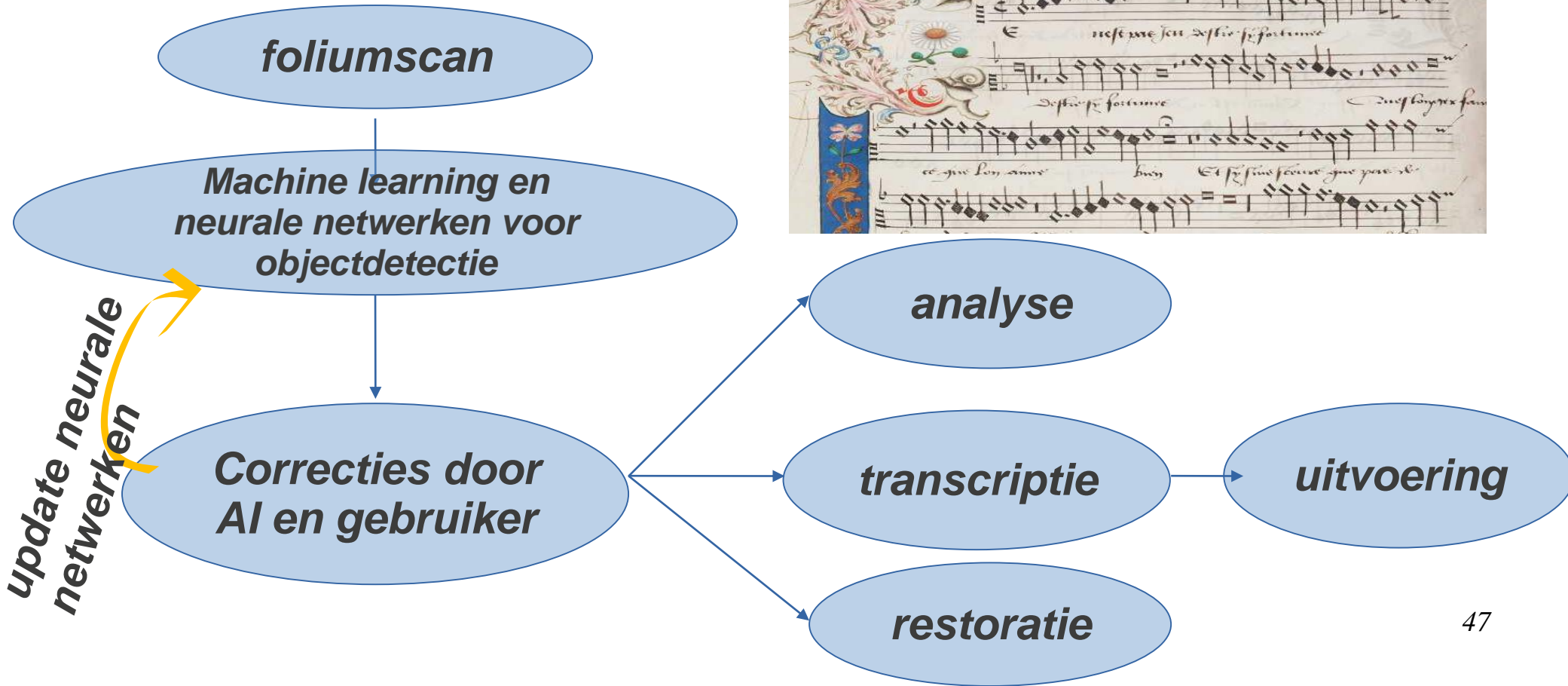


Nootduraties hangen af van omliggende noten



*Ligatuur* : meerdere noten voor één lettergreep

# Ontsluiten van polyfonie: pipeline



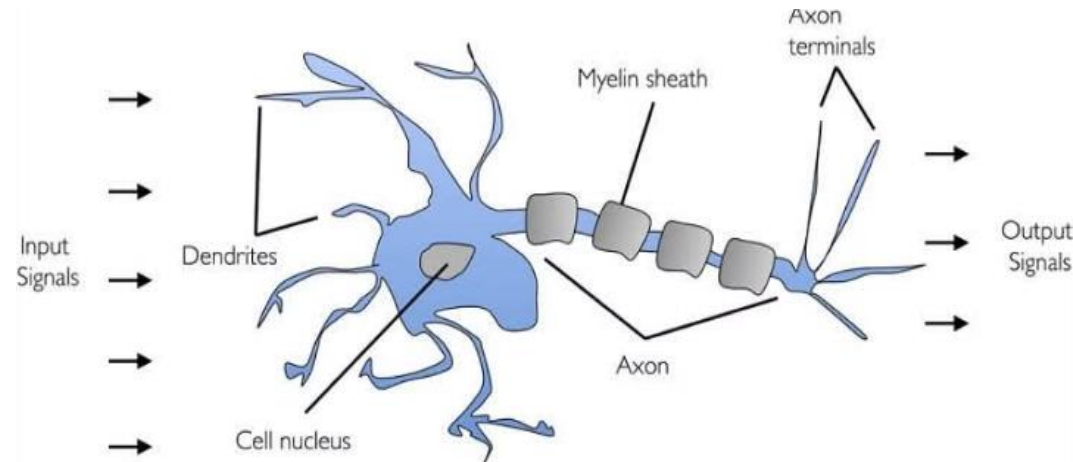
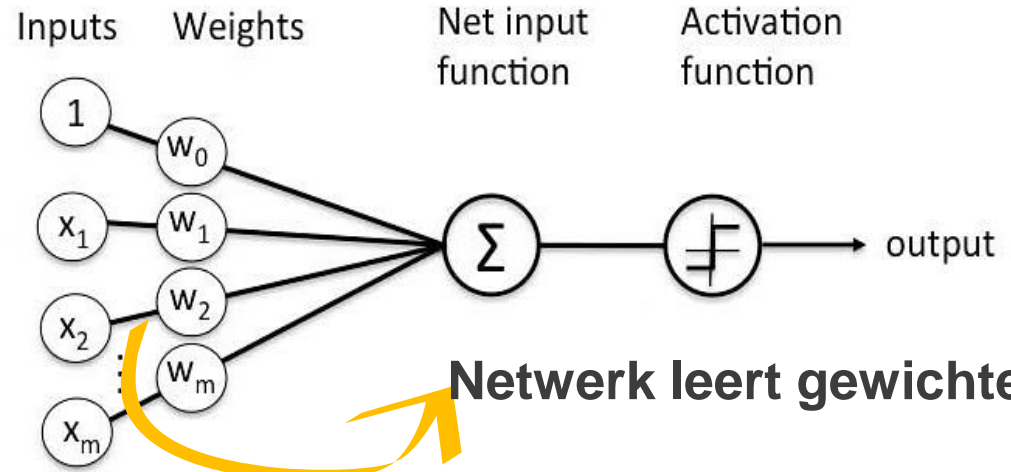
# Ontsluiten van polyfonie

## Het wiskundig neuron:

Bouwsteen van neurale netwerken;  
ontleend aan het biologische neuron,  
bouwsteen van het brein

## werking

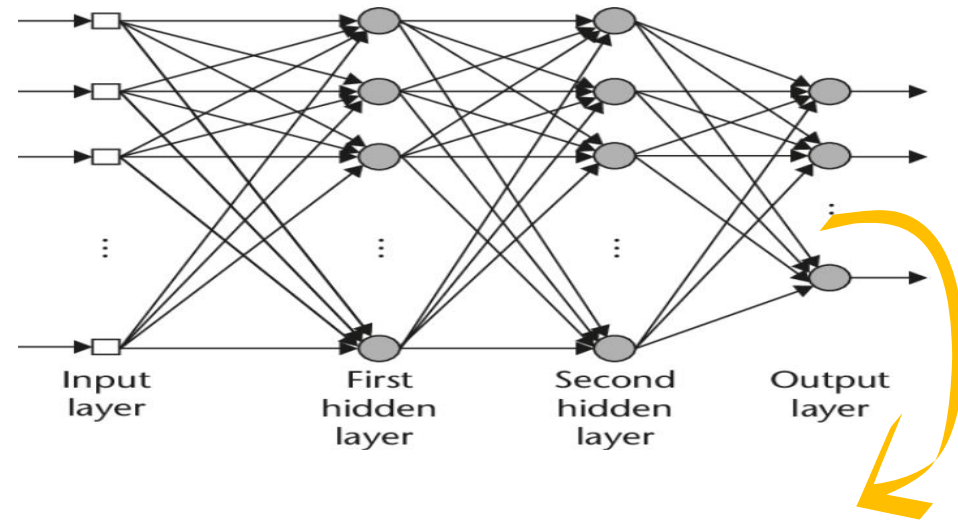
- Ontvangt input van andere neuronen
- Hecht verschillend belang/gewicht aan verschillende inputs
- Sommatie gewogen inputsignalen
- Doorlaten (of niet) van gemoduleerd signaal





# Ontsluiten van polyfonie

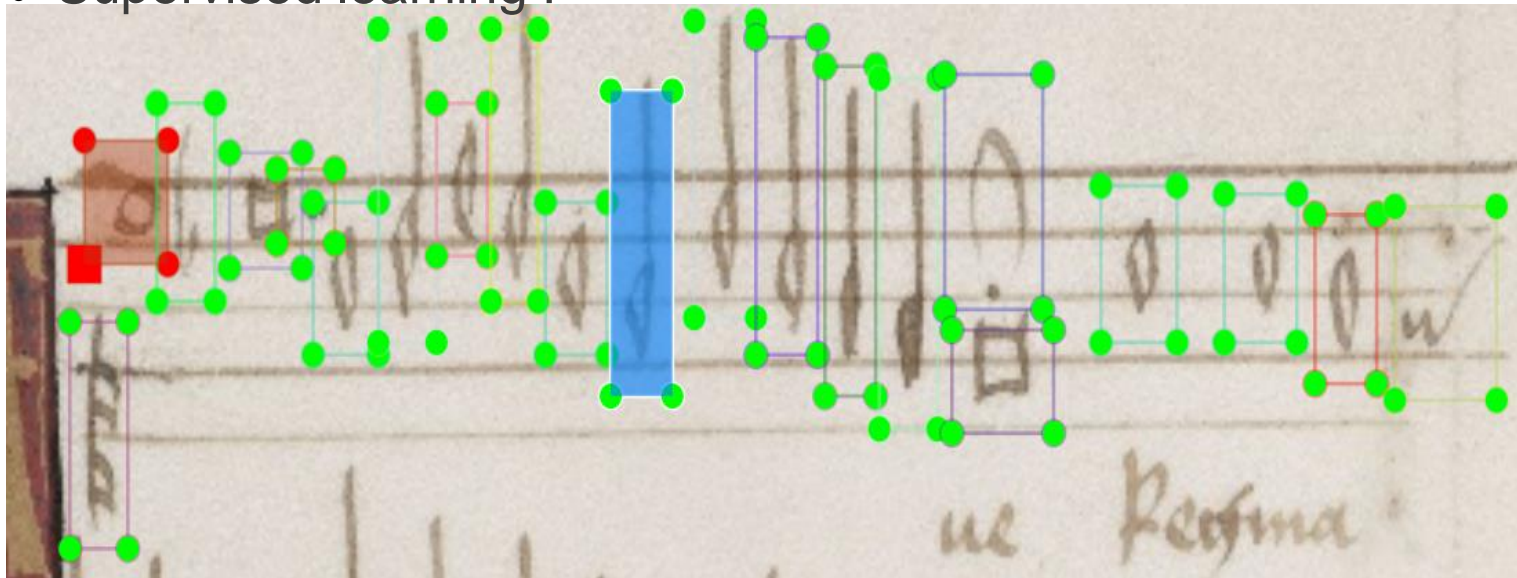
- Verschillende neuronen, in diepe neurale netwerken zijn nodig om visuele patronen te leren (trainen).
- Deze geleerde patronen worden dan gematcht met regio's op afbeelding (convolutie)



Output layer met waarschijnlijkheden voor alle mogelijke classificatiecategorieën

# Ontsluiten van polyfonie

- Neurale netwerken hebben véél data nodig om van te leren.
- Expert annotateert handmatig rechthoekjes met duratie en toonhoogte rond symbolen
- Alamire databank, **IDEM**, heeft 80 000 juiste symbolen in 250 folia om van te leren
- Supervised learning !



<input checked="" type="checkbox"/>	c clef_11
<input checked="" type="checkbox"/>	Pmin
<input checked="" type="checkbox"/>	flat_5
<input checked="" type="checkbox"/>	breve_4
<input checked="" type="checkbox"/>	point
<input checked="" type="checkbox"/>	semibreve_6
<input checked="" type="checkbox"/>	minim_5
<input checked="" type="checkbox"/>	semibreve_3
<input checked="" type="checkbox"/>	minim_4
<input checked="" type="checkbox"/>	semibreve_6
<input checked="" type="checkbox"/>	minim_7
<input checked="" type="checkbox"/>	minim_5
<input checked="" type="checkbox"/>	minim_6
<input checked="" type="checkbox"/>	semiminim_7
<input checked="" type="checkbox"/>	semiminim_8
<input checked="" type="checkbox"/>	fermata
<input checked="" type="checkbox"/>	breve_9
<input checked="" type="checkbox"/>	semibreve_6
<input checked="" type="checkbox"/>	semibreve_6
<input checked="" type="checkbox"/>	semibreve_7
<input checked="" type="checkbox"/>	custos_8

# Ontsluiten van polyfonie

## Enkele moeilijkheden

Inkdoorsijpeling en degradatie zorgen voor *low confidence* matches of verkeerde matches door het neurale netwerk



Het neurale netwerk moet in staat zijn om notatie in verschillende handschriften te herkennen, ook stijlen die niet gezien werden tijdens het trainen

# Ontsluiten van polyfonie

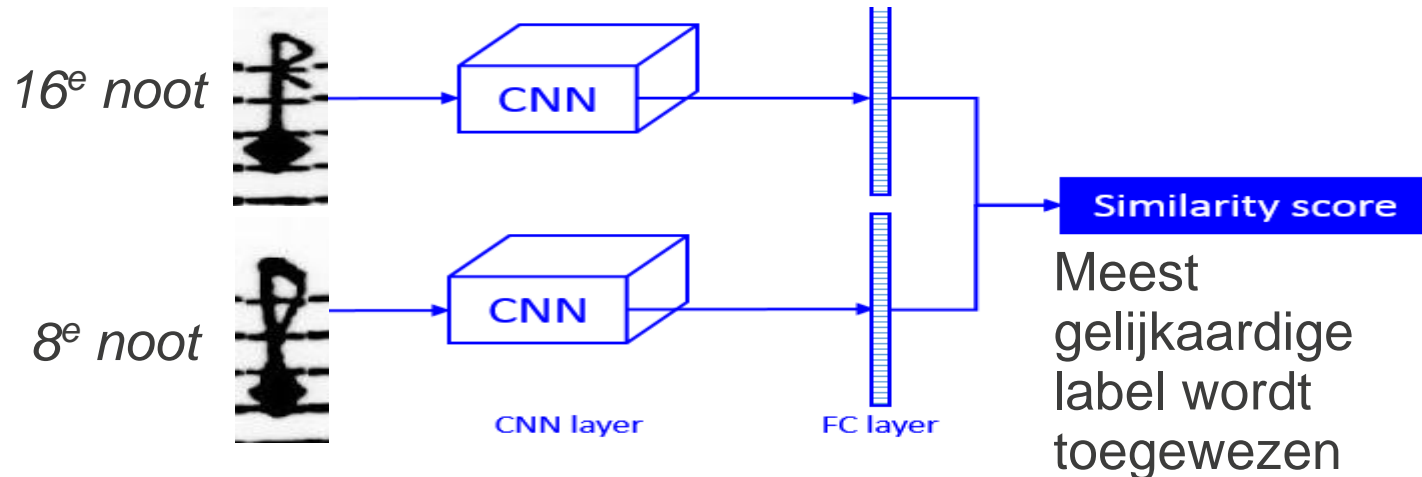
## Enkele moeilijkheden

Sommige zeldzame symbolen lijken op elkaar en worden *meestal*, maar niet altijd juist onderscheiden door het initiële objectdetectie netwerk



## Siamese netwerken

Symbolen waarover het netwerk onzeker is, worden vergeleken met hoog waarschijnlijke symbolen uit het folium en symbolen uit databank





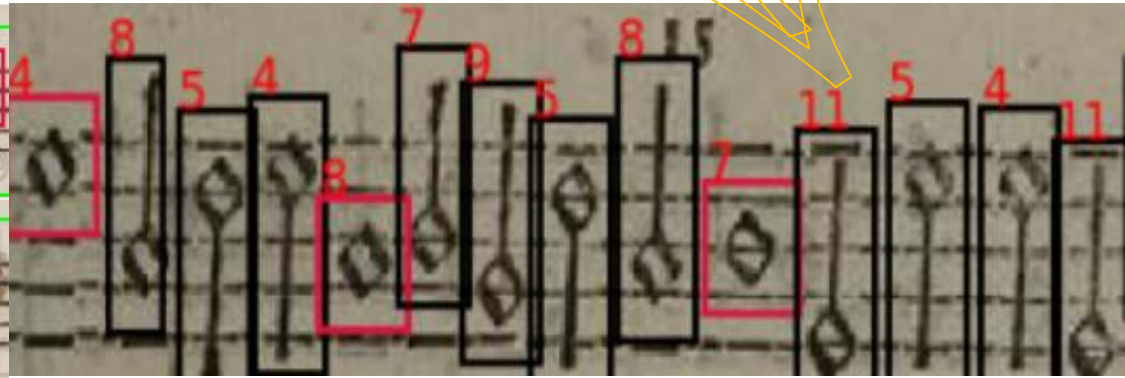
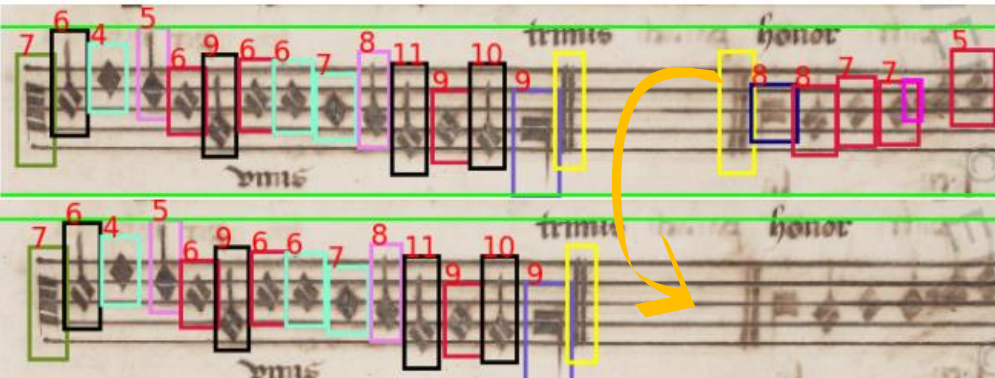
# Ontsluiten van polyfonie

Noodzaak van correcties door AI met logica na objectdetectie

Vals positieven ten gevolge van  
Inktdoorsijpeling. Corrigeerbaar door logica,  
want na dubbele maatstreep

Foutievelijk gedetecteerde nootpositie  
corrigeerbaar door taal/tijdreeksmodel  
met logica, want septiemsprong

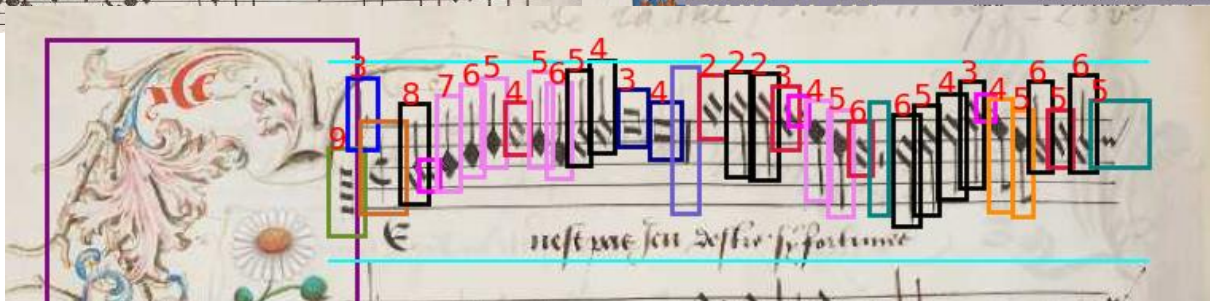
$E4/D4 \approx 15/8$   
eerder  
dissonant



# Ontsluiten van polyfonie



Van  
manuscript  
tot transcriptie  
met AI



Symbol-  
detectie met  
 $\approx 99\%$   
accuraatheid

# Ontsluiten van polyfonie

**State of the art** resultaten voor muziekherkenning door de **kracht van het ensemble**: objectdetectie + taalmodel + siamees netwerk + musicologische heuristieken

## To do:

Muziek is innig verwoven met het dagelijkse leven, bronnen bevatten naast muzieknotatie:

illuminaties

heraldiek

tekst

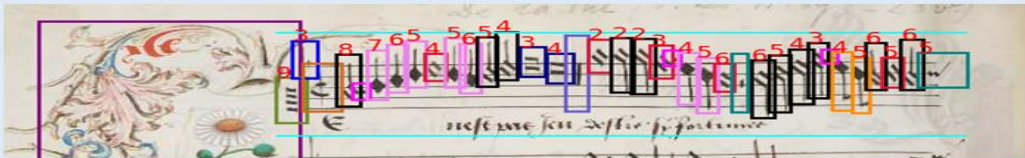


**Nood aan multi-disciplinaire benadering....**





# Flanders AI 2.0



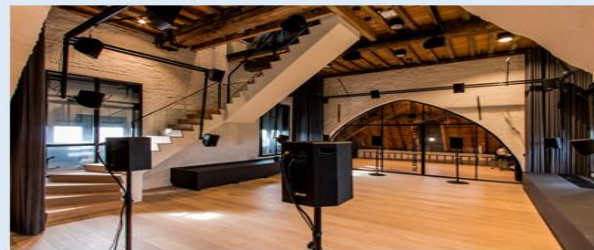
Unlocking Polyphony



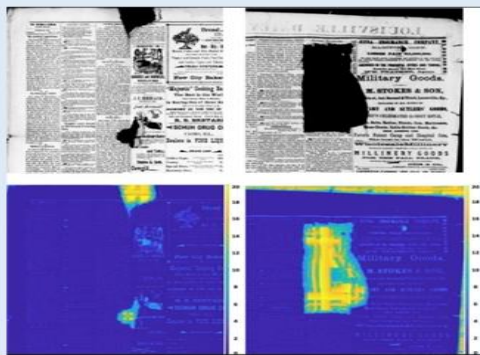
Digital Heritage



Unlocking GLAM Collections



Acoustics as intangible cultural heritage



Unlocking Cultural Heritage Artefacts



The old University



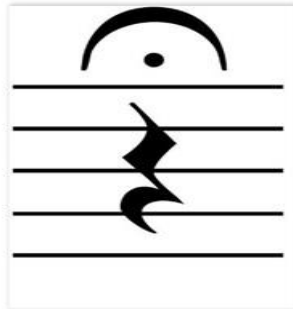
Curation of spoken documents



*Multi-, inter- en cross-disciplinariteit:*

*wiskunde, fysica, AI, geschiedenis,  
ingenieurswetenschappen...*

*= de toekomst van alle wetenschappen !*



# Referenties

*Slide 1: Mechels koorboek, [exponanza.be](http://exponanza.be)*

*Slide 2: [biologielessen.nl](http://biologielessen.nl)*

*Slide 3: [wikipedia.org/wiki/String\\_vibration](http://wikipedia.org/wiki/String_vibration)*

*Slide 6: [galileo.phys.virginia.edu/classes/152.mf1i.spring02/AnalyzingWaves.htm](http://galileo.phys.virginia.edu/classes/152.mf1i.spring02/AnalyzingWaves.htm)*

*Slide 8: [pianostreet.com/smf/index.php?topic=64233.msg681347#msg681347](http://pianostreet.com/smf/index.php?topic=64233.msg681347#msg681347)*

*Slide 8: [istockphoto.com/fr/photos/guitar-tuner](http://istockphoto.com/fr/photos/guitar-tuner) , [4strings.be/nl/snarengids/help-welke-snaren-heb-ik-nodig/](http://4strings.be/nl/snarengids/help-welke-snaren-heb-ik-nodig/)*

*Slide 10: [de.wikipedia.org/wiki/Sinus\\_und\\_Kosinus](http://de.wikipedia.org/wiki/Sinus_und_Kosinus)*

*Slide 12: [toptal.com/algorithms/shazam-it-music-processing-fingerprinting-and-recognition](http://toptal.com/algorithms/shazam-it-music-processing-fingerprinting-and-recognition)*

*Slide 13,14: [projectrhea.org/rhea/index.php/Fourier\\_analysis\\_in\\_Music](http://projectrhea.org/rhea/index.php/Fourier_analysis_in_Music)*

*Slide 18: [yoair.com/blog/the-significance-of-greek-muses-in-ancient-and-modern-times/](http://yoair.com/blog/the-significance-of-greek-muses-in-ancient-and-modern-times/)*

*Slide 19: [wikipedia.org/wiki/Sonata\\_form](http://wikipedia.org/wiki/Sonata_form)*

*Slide 20: [gregobase.selapa.net/chant.php?id=6043](http://gregobase.selapa.net/chant.php?id=6043)*

*Slide 20: [youtube.com/@EarlyMusicSources](http://youtube.com/@EarlyMusicSources)*

# Referenties

*Slide 21: [en.wikipedia.org/wiki/Cadence](https://en.wikipedia.org/wiki/Cadence) , [en.wikipedia.org/wiki/I-V-vi-IV\\_progression](https://en.wikipedia.org/wiki/I-V-vi-IV_progression)*

*Slide 24: [en.wikipedia.org/wiki/Sampling\\_%28signal\\_processing%29](https://en.wikipedia.org/wiki/Sampling_%28signal_processing%29)*

*Slide 25: [en.wikipedia.org/wiki/Spectrogram](https://en.wikipedia.org/wiki/Spectrogram)*

*Slide 27-31: A Wang, An Industrial-Strength Audio Search Algorithm, ISMIR 2003  
[www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf](http://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf)*

*Slide 33-40: <https://tvanwate.github.io/vaia.pdf>*

*Slide 43-45: Alamari Foundatiton, [alamirefoundation.org](http://alamirefoundation.org)*

*Slide 46: M. Thomae, Automatic soring up of mensural music using perfect mensurations*

*Slide 48: [simplilearn.com/tutorials/deep-learning-tutorial/perceptron](https://simplilearn.com/tutorials/deep-learning-tutorial/perceptron)*

*Slide 49: [stackoverflow.com/questions/42883547/intuitive-understanding-of-1d-2d-and-3d-convolutions-in-convolutional-neural-n](https://stackoverflow.com/questions/42883547/intuitive-understanding-of-1d-2d-and-3d-convolutions-in-convolutional-neural-n)*

*Slide 50: [deviantart.com/sklavenbrause/art/We-are-siamese-if-you-please-109129425](https://deviantart.com/sklavenbrause/art/We-are-siamese-if-you-please-109129425)*

*Slide 55: T. Tuytelaars, H. van Hamme, T. van Waterschoot, Alamire Foundation, A. Dooms, F. Truyen, Y. Uyttendaele*