A brief introduction the the physics of taps, clangs, crashes, scrapes, crackles, flutters, rattles, pops, sloshes, splashes, gurgles, whooshes and cracks

James Traer

McDermott Computational Audition Laboratory
Brain and Cognitive Sciences Department
MIT

June 17, 2015

Outline

Introduction

1D Model

2D and 3D Models

Speech production model

Statistics of the forcing function

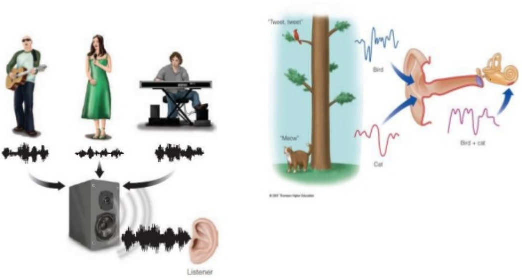
A whirlwind tour of environmental sounds

Conclusion

Sound conveys information, from which scenes and actions can be inferred

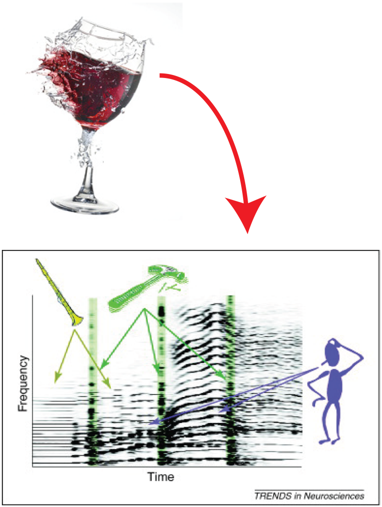
Auditory Scene Analysis

Perceptually Organizing Sounds in the Environment



Example 1

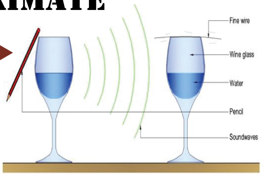
Physical properties of materials determine the properties of sound pressure waves



- How does the brain interpret the structure of sound?
- How does the physical environment constrain the structure of sound?
- How does the brain infer the physical environment from sound?

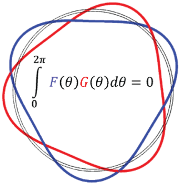
Simple physical models can explain many properties of complicated sound sources

1. APPROXIMATE



$$\ddot{\mathbf{x}} = \frac{\mathbf{F}}{m} \qquad \frac{\mathrm{d}E}{\mathrm{d}t} = 0 \qquad \frac{\mathrm{d}(m\dot{\mathbf{x}})}{\mathrm{d}t} = 0$$

2. SOLVE



3. PRAY

Overview

- 1D oscillators: a painfully detailed analysis of a simple idealized string
  - Wave equation has oscillitory solutions
  - Normal modes predict harmonic spectra
  - Onsets remain difficult
  - Sound radiation is directional
  - Multiple models can be used for similar systems
  - Non-linearities are hard
- 2D and 3D oscillators: same equation, harder algebra, same general results
- Speech generation
- Modelling the statistics of the energetic forcing
- A whirlwind tour through the physics underlying common environmental sounds

Outline

Introduction

1D Model

2D and 3D Models

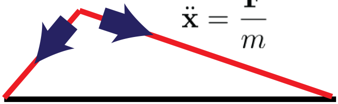
Speech production model

Statistics of the forcing function

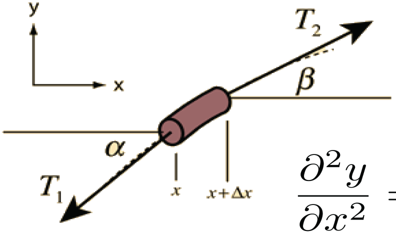
A whirlwind tour of environmental sounds

Conclusion

A 1D string can be (crudely) modeled by a wave equation



$$\ddot{\mathbf{x}} = \frac{\mathbf{F}}{m}$$



ρ = mass per unit length

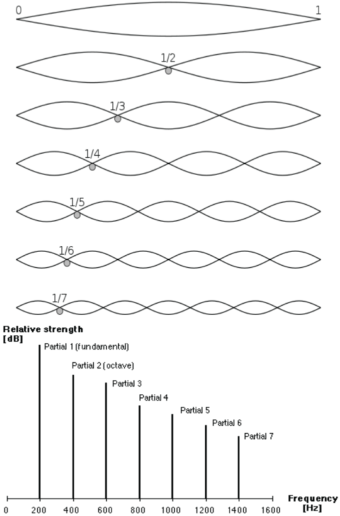
$$\frac{\partial^2 y}{\partial x^2} = \frac{\rho}{T}\frac{\partial^2 y}{\partial t^2}$$

- Solution:

$$\begin{aligned} y &= Ae^{\imath k(x-ct)} \\ c &= \sqrt{\frac{T}{\rho}} \end{aligned}$$

- oscillatory solutions
- Frequency and wavelength are related by wave speed

The boundary conditions impose harmonic solutions (Normal Modes)

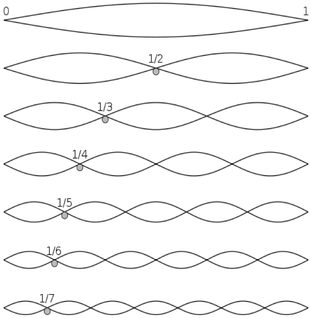


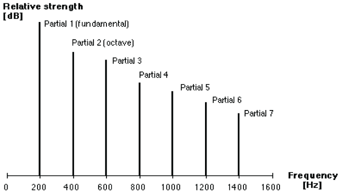
- Solution:

$$y = Ae^{\imath k(x-ct)}$$

- two fixed ends: $y(0) = 0; \qquad y(L) = 0$
- discretizes spatial solutions $k = \frac{2\pi}{L}$
- due to intrinsic relationship between spatial and temporal modes, this discretizes temporal solutions too
  Some example strings: Guitar; Bass

If oscillation is forced the string acts as a filter

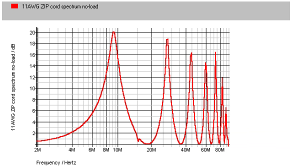




- $$F(t) = \frac{\rho}{T}\frac{\partial^2 y}{\partial t^2} - \frac{\partial^2 y}{\partial x^2}$$

- Solution:

$$y = Ae^{\imath k(x-ct)} + \alpha F(t)$$



- Example: Lab demo

Real world strings are more complicated but share similarities to the cartoon model

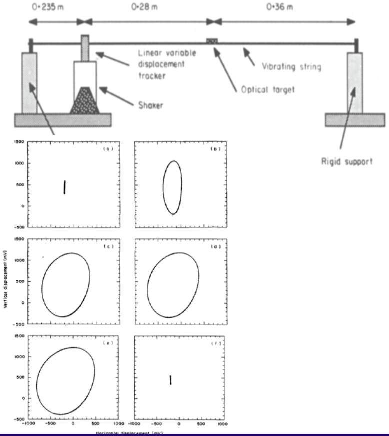
We have made a number of approximations

- No friction
- => exponential decay term to each mode: $y = Ae^{-bt}e^{\imath k(x-ct)}$
- A delta function initial disturbance
- => real impulses have a finite spectrum
- None-modal oscillations exist transiently
- => attacks are broadband
- We have linearized twice
- (1) Constant tension; (2) small angle
- (1) Tension changes allow longitudinal waves (Example); (2) the small angle approximation is often violated during attacks

With different assumptions the same system can be described by multiple models

Non-linearities allow energy to be transferred between different frequency modes

$$\ddot{\mathbf{x}} = \left(1 + \Gamma \int_0^L \dot{\mathbf{x}}^2 \mathrm{d}x\right) \nabla^2 \mathbf{x}$$

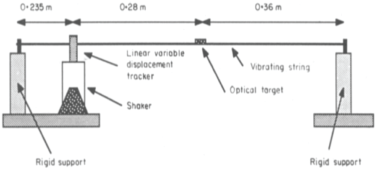


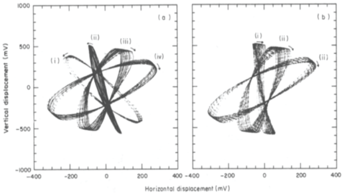
- Non-linear equations are fundamentally more difficult
- The solutions are not necessarily stable
- Energy can be exchanged between different modes (i.e ψ_y (f_1, λ_1) => ϕ_x (f_1, λ_2) => ψ_y (f_2, λ_2))

O'Reilly and Holmes, J. Sound Vibration, 1992

Non-linearities allow energy to be transferred between different frequency modes

$$\ddot{\mathbf{x}} = \left(1 + \Gamma \int_0^L \dot{\mathbf{x}}^2 \mathrm{d}x\right) \nabla^2 \mathbf{x}$$



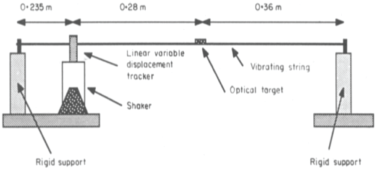


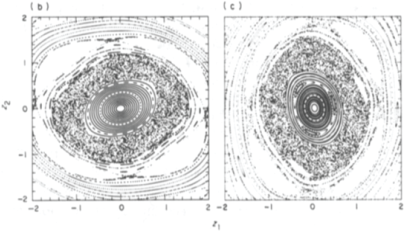
- Non-linear equations are fundamentally more difficult
- The solutions are not necessarily stable
- Energy can be exchanged between different modes (i.e ψ_y (f_1, λ_1) => ϕ_x (f_1, λ_2) => ψ_y (f_2, λ_2))

O'Reilly and Holmes, J. Sound Vibration, 1992

Non-linearities allow energy to be transferred between different frequency modes

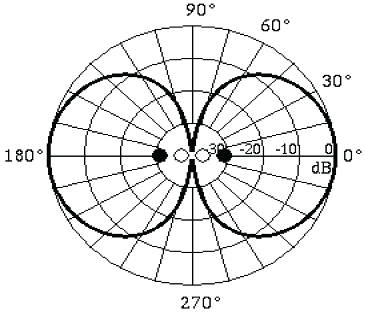
$$\ddot{\mathbf{x}} = \left(1 + \Gamma \int_0^L \dot{\mathbf{x}}^2 \mathrm{d}x\right) \nabla^2 \mathbf{x}$$





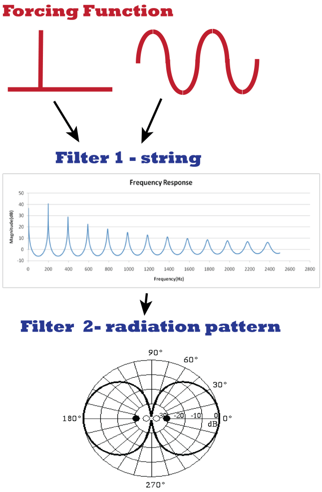
- Non-linear equations are fundamentally more difficult
- The solutions are not necessarily stable
- Energy can be exchanged between different modes (i.e ψ_y (f_1, λ_1) => ϕ_x (f_1, λ_2) => ψ_y (f_2, λ_2))

The geometry of the sound source effects the radiation pattern



- To reach the ear, sound must couple to the air
- Model: solve the 3D wave-eqn with 1D forcing function
- Radiation is not emitted uniformly
- Sound depends upon spatial relationship between source and the listener

Sound from a 1D string can therefore be modelled as a forcing function and nested filters



- Some forcing function $F(t, \mathbf{x})$ sets the string in motion
- The resonances (modes) of the string filter the vibration
- The radiative pattern filters this spatially
- The environment filters yet again (reverb)
- Each of these steps can be modelled by (usually linear) filters:
  $s(t) = F(t, \mathbf{x}) * h_{string} * h_{radiation} * h_{reverb}$
- The spectro-temporal properties of each filter are determined by geometry and material properties
- Many important features of the sound are due to the forcing $F(t, \mathbf{x})$ Example

Outline

Introduction

1D Model

2D and 3D Models

Speech production model

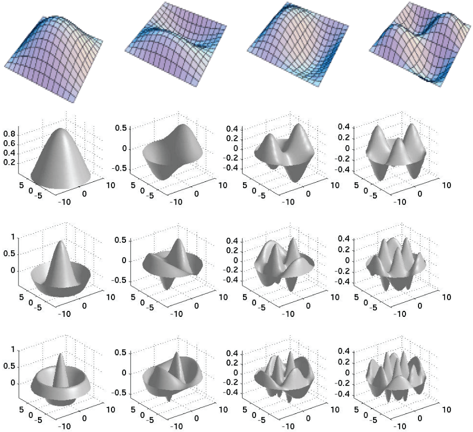
Statistics of the forcing function

A whirlwind tour of environmental sounds

Conclusion

A 2D membrane can be modeled by the same wave and radiation equations as a string

$$\frac{\partial^2 U}{\partial t^2} = a^2\left(\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2}\right)$$



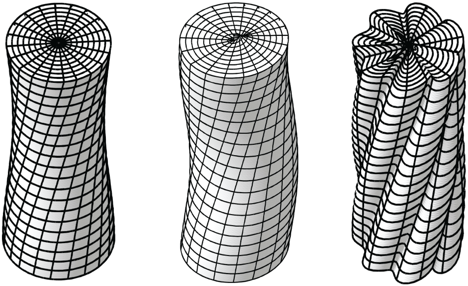
- As with the 1D model the membrane has resonant modes
- They are not necessarily harmonic

Ex: Driven Chladni Plate

Ex: Bowed Chladni Plate

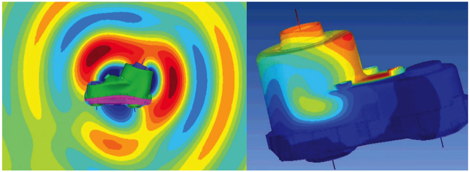
3D blocks and plates are amenable to the same models

$$\rho \frac{\partial v_x}{\partial t} = \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + \frac{\partial \sigma_{xz}}{\partial z} + f_x$$
$$\rho \frac{\partial v_y}{\partial t} = \frac{\partial \sigma_{yx}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + \frac{\partial \sigma_{yz}}{\partial z} + f_y$$
$$\rho \frac{\partial v_z}{\partial t} = \frac{\partial \sigma_{zx}}{\partial x} + \frac{\partial \sigma_{zy}}{\partial y} + \frac{\partial \sigma_{zz}}{\partial z} + f_z. \quad (1)$$



- Geometry enforces normal modes
- They may not be harmonic
- Multiple harmonic series associated with shearing and pressure waves
- Damping induces exponential decay
- The radiation pattern is likely to be complicated
- Non-linearities are not uncommon (thin shells)

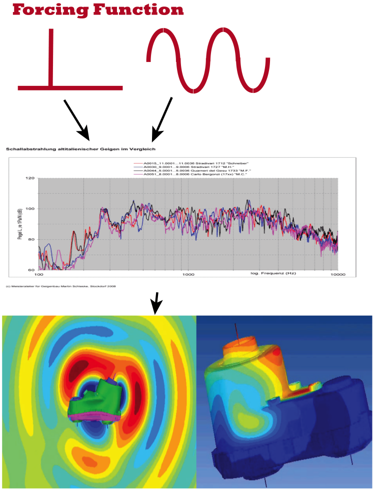
3D blocks and plates are amenable to the same models



$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = F(t, \mathbf{x})$$

- Geometry enforces normal modes
- They may not be harmonic
- Multiple harmonic series associated with shearing and pressure waves
- Damping induces exponential decay
- The radiation pattern is likely to be complicated
- Non-linearities are not uncommon (thin shells)

The logical framework of the 1D model still applies, only the algebra is harder



- Finite-element analysis allows the wave equation to be solved in arbitrarily complicated shapes
- The nested filter model still applies:
  $s(t) = F(t, \mathbf{x}) *$
  $h_{3D-shape} * h_{radiation} * h_{reverb}$
- Example (9:38-12:33)
- Even non-linear systems can be hacked (27:28-28:16; 32:49-33:14)

Crude features of the model are robust to a variety of situations

- Size + Density => Bandwidth
- Geometric regularity => Harmonicity
- Density and elasticity => Damping
- Flexing => Spectro-temporal statistics
- Shape, position, distance + motion => alter spectro-temporal statistics

Outline

Introduction

1D Model

2D and 3D Models

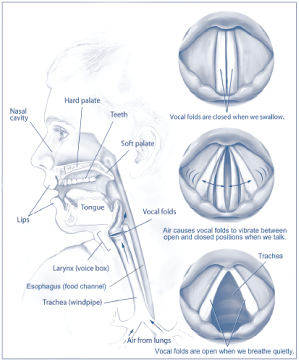
Speech production model

Statistics of the forcing function

A whirlwind tour of environmental sounds

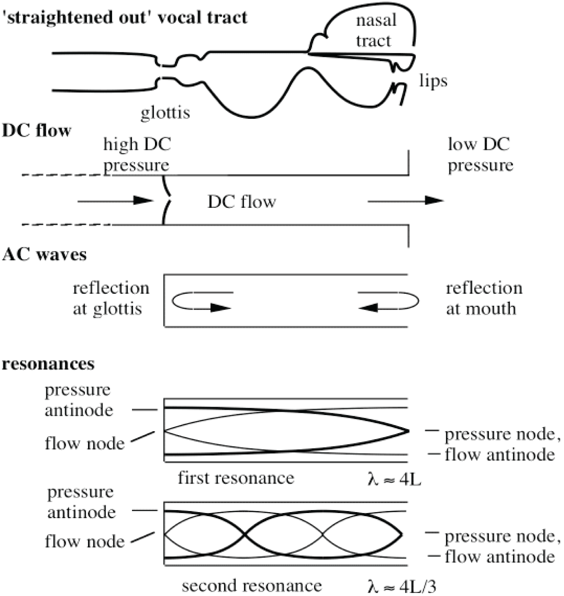
Conclusion

Speech production is modelled by a harmonic source (vocal chords) and a resonant filter (vocal tract)



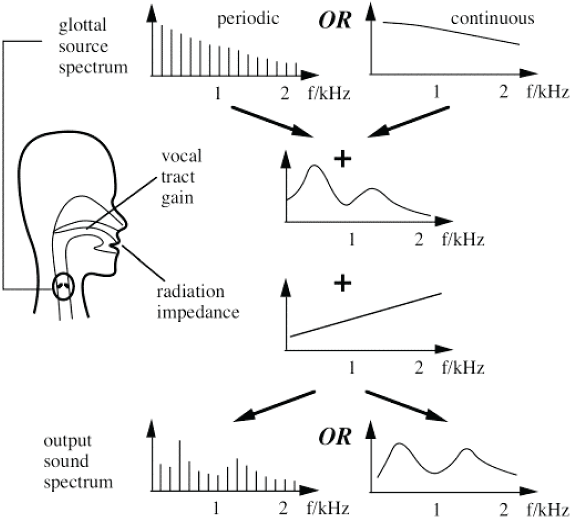
- The nested filter model still applies: $s(t) = F(t, \mathbf{x}) * h_{VocalChords} * h_{VocalTract} * h_{radiation} * h_{reverb}$
- However, the filters are time-varying...
- ... and this variation provides much of the structure that we actually care about in speech

Speech production is modelled by a harmonic source (vocal chords) and a resonant filter (vocal tract)



- The nested filter model still applies: $s(t) = F(t, \mathbf{x}) * h_{VocalChords} * h_{VocalTract} * h_{radiation} * h_{reverb}$
- However, the filters are time-varying...
- ... and this variation provides much of the structure that we actually care about in speech

Speech production is modelled by a harmonic source (vocal chords) and a resonant filter (vocal tract)



- The nested filter model still applies: $s(t) = F(t, \mathbf{x}) * h_{VocalChords} * h_{VocalTract} * h_{radiation} * h_{reverb}$
- However, the filters are time-varying...
- ... and this variation provides much of the structure that we actually care about in speech

Outline

Introduction

1D Model

2D and 3D Models

Speech production model

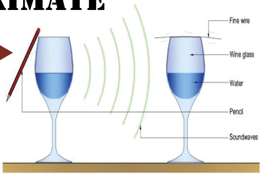
Statistics of the forcing function

A whirlwind tour of environmental sounds

Conclusion

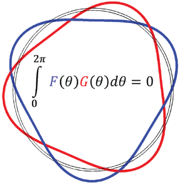
Recap: Simple physical models can explain many properties of complicated sound sources

1. APPROXIMATE



$$\ddot{\mathbf{x}} = \frac{\mathbf{F}}{m} \qquad \frac{\mathrm{d}E}{\mathrm{d}t} = 0 \qquad \frac{\mathrm{d}(m\dot{\mathbf{x}})}{\mathrm{d}t} = 0$$

2. SOLVE



3. PRAY

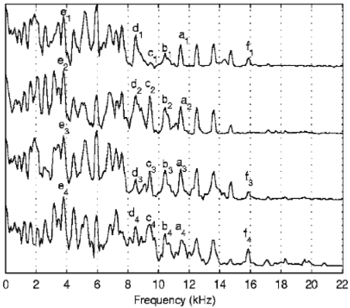
Recap: The algebra is tedious but tractable

$$s(t) = F(t, \mathbf{x}) * f_1 * f_2 * \ldots * f_{reverb}$$

The filter properties are determined by physical properties:

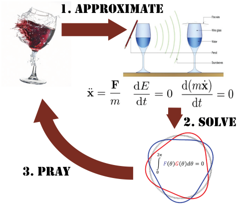
- Size + Density => Bandwidth
- Geometric regularity => Harmonicity
- Density and elasticity => Damping
- Flexing => Spectro-temporal statistics
- Shape, position, distance + motion => alter spectro-temporal statistics
- Example: 32:30

Structure in the Forcing Function results in structure in the sound



- Sound may depend sensitively on the forcing function $F(t, \mathbf{x})$
- Filters may be easy to model: $s(t) =$ $F(t, \mathbf{x}) * h_1 * h_2 * h_{radiation} * h_{reverb}$
- But realistic synthesis may require modelling the forcing fuction as well $F(t)$
- This is relatively unexplored by acousticians...
- ... but not by everyone. Example (12:33-13:14; 14:24-14:58; 15:27-16:47; 41:12-42:43)

Spoiler-alert: Sound can be well modelled by acoustical physics and regular physics



- Cartoon models of the world predict:
  - Size + Density => Bandwidth
  - Geometric regularity => Harmonicity
  - Density and elasticity => Damping
  - Flexing => Spectro-temporal statistics
  - Shape, position, distance + motion => alter spectro-temporal statistics
- The statistics of the forcing function $F(t, \mathbf{x})$ affect the statistics of the sound
- Realistic sound synthesis requires modelling object and filter dynamics in addition to acoustic properties
- The filters are largely solved (k-wave toolbox for matlab)
- The dynamics of this forcing function are just more physics. And are often tractable.

Outline

Introduction

1D Model

2D and 3D Models

Speech production model

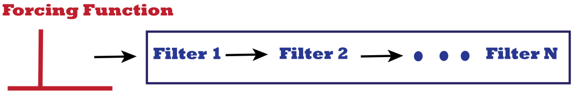
Statistics of the forcing function

A whirlwind tour of environmental sounds

Conclusion

bangs , crashes , clangs , claps , taps , knocks and rings

Impact driven resonant bodies



$$s(t) = f_N\left(\ldots f_2\left(f_1(F(t))\right)\ldots\right)$$

Filters can be

- Spectrally sharp or broad (geometry, density elasticity)
- Bandlimited (Size, density)
- harmonic or inharmonic (geometry)
- Ringing or damped (density, elasticity)
- Linear or Non-linear (flexibility)

Rattles , buzzes and hums

Periodic impacted driven resonant bodies

$$s(t) = f_N\left(\ldots f_2\left(f_1(F(t))\right)\ldots\right)$$

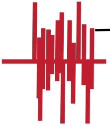
- Filters are identical to the impulse cases
- Statistics of impacts are important

- Filters are identical to the impulse cases
- Only the statistics of the forcing function have changed

Scrapes , grates , rolls , squeaks , creaks , squeals , screeches , renders , rustles , brushes , sweeps and buzzes

Frictionally driven resonant bodies

Forcing Function



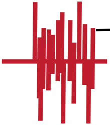
Filter 1 → Filter 2 → • • • Filter N

$$s(t) = f_N\left(\ldots f_2\left(f_1(F(t))\right)\ldots\right)$$

Hisses , whistles , toots , whooshes , rushes , gushes , flutters and roars

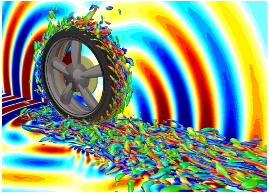
Turbulent air flow

Forcing Function



Filter 1 → **Filter 2** → • • • **Filter N**

$$s(t) = f_N\left(\ldots f_2\left(f_1(F(t))\right)\ldots\right)$$



- Source is white noise-like
- Often coupled to a resonating body
- Typically the filtering is less dramatic than in friction sounds – hence they sound quite different Example (37:48)

Gurgles , drops , bubbles , sloshes , bloops , laps , crashes and splashes

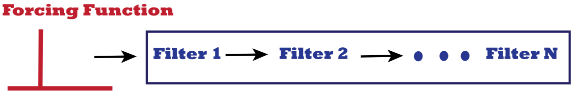
Cavitation

$$s(t) = f_N\left(\ldots f_2\left(f_1(F(t))\right)\ldots\right)$$

- Source is dependant upon statistics of bubbles
- Sound is filtered in rapidly changing resonant cavities Example (37:22)

Cracks , claps and booms

Acceleration driven shock waves

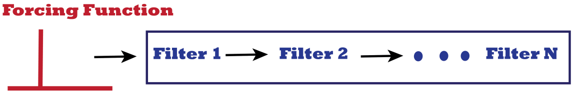


$$s(t) = f_N\left(\ldots f_2\left(f_1(F(t))\right)\ldots\right)$$

- Impulse is usually louder than the resonant modes
- (Except in the case of thunder)
- Many pedestrian sounds have an impulsive component: Example (23:07)

Shattering , splintering , splitting , bursting and crushing

Deforming resonant oscillators



$$s(t) = f_N\left(\ldots f_2\left(f_1(F(t))\right)\ldots\right)$$

- The number of resonators and their normal modes change with time

Outline

Introduction

1D Model

2D and 3D Models

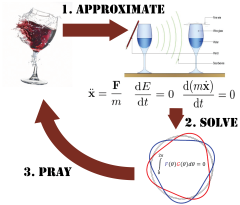
Speech production model

Statistics of the forcing function

A whirlwind tour of environmental sounds

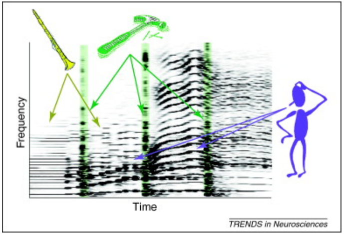
Conclusion

Real-world sound sources are complicated to model, but simplified models allow realistic sound synthesis



- Cartoon models of the world predict:
  - Size + Density => Bandwidth
  - Geometric regularity => Harmonicity
  - Density and elasticity => Damping
  - Flexing => Spectro-temporal statistics
  - Shape, position, distance + motion => alter spectro-temporal statistics
- Realistic sound synthesis requires modelling object and filter dynamics in addition to acoustic properties
- This is just more physics and can be crunched to synthesize quite realistic sounds (with patience and a lot of CPU power)

Final thoughts...



- The statistics of sounds are filtered versions of the statistics of motion
- The properties of the filters are determined by the physical properties of everyday objects
- Both the filters and the motion are largely solved problems
- ... but only very recently have they are rarely combined to synthesize compelling sounds
- The inference problem remains to be solved (by scientists):

$F(t, \mathbf{x}) = ?$ $\qquad$ $f_{resonator} = ?$

The final word to Lord Rayleigh....

1. THE sensation of sound is a thing *sui generis*, not comparable with any of our other sensations. No one can express the relation between a sound and a colour or a smell. Directly or indirectly, all questions connected with this subject must come for decision to the ear, as the organ of hearing; and from it there can be no appeal. But we are not therefore to infer that all acoustical investigations are conducted with the unassisted ear. When once we have discovered the physical phenomena which constitute the foundation of sound, our explorations are in great measure transferred to another field lying within the dominion of the principles of Mechanics. Important laws are in this way arrived at, to which the sensations of the ear cannot but conform.