Advanced Animatronics Voice ! v0.1

Flüüfff – 12/11/2022

Floere T. Pillowcase, Devourer of Automobiles

floere@robocow.be

What is this Talk About ?

The surprising **complexity** of **hearing and speaking** while in suit

- Basic **recording studio** techniques
- Advanced signal processing **algorithms** that make it all possible

This is all about **sound reproduction** and **sound processing**!

(Incidentally, you may also get some pointers for your live streaming set-up. And learn a few things about how modern cellphones & video conferencing softwares do their thing)

Content

- **The Goal**
- **The Concept**
- **Why is this so Hard?**
- **Wrapping-Up**
- **Demonstration**

The Goal: Puppet Without the Strings

- **Your character transparently driven by your acting**

- **Clear voice projection**

- **Live audience interaction**

- **Everything self-contained in the costume**

- **Comfortable**

- **Affordable(-ish)**

RiverDai
@ParksbyRiver · Follow

Shanghai Disneyland reveals an Officer Clawhauser animatronic in the upcoming #Zootopia attraction ahead of #D23Expo Day3.

Watch on Twitter

2:06 PM · Sep 11, 2022

♥ 25K   💬 Reply   🔗 Copy link

Read 224 replies

What ! Where's Radula Castion ?

Testing Elleven – Radula Castion – https://www.youtube.com/watch?v=EKWqEGFbLAg

Voice System Capabilities

- **Easy enough to implement by hobbyists and small companies**

  - Mainly off-the-shelf parts

  - The remainder must be maker friendly
    (hand tools – 3D-printer – laser cutter)

- **Makes the suit head "transparent" for sound**

  - Hear what happens around you

  - Clear voice to the audience

  - Can displace the creature's head from yours

    - Animated jaw + speaker/mic in displaced head!

Kazul – "Hogger"
https://www.youtube.com/c/KazplayVideos/videos
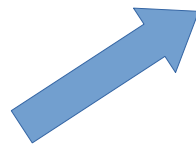
The Character Shop – "Pedroso"
http://ww.character-shop.com/pedroso.html

Content

- **The Goal**

- **The Concept**

- **Why is this so Hard?**

  - Interference Effect

  - Larsen Effect

  - Close-Talking, Directional Microphone Issues

  - Suit Head and Loudspeaker Resonances

  - Head-Related Transfer Functions

- **Wrapping-Up**
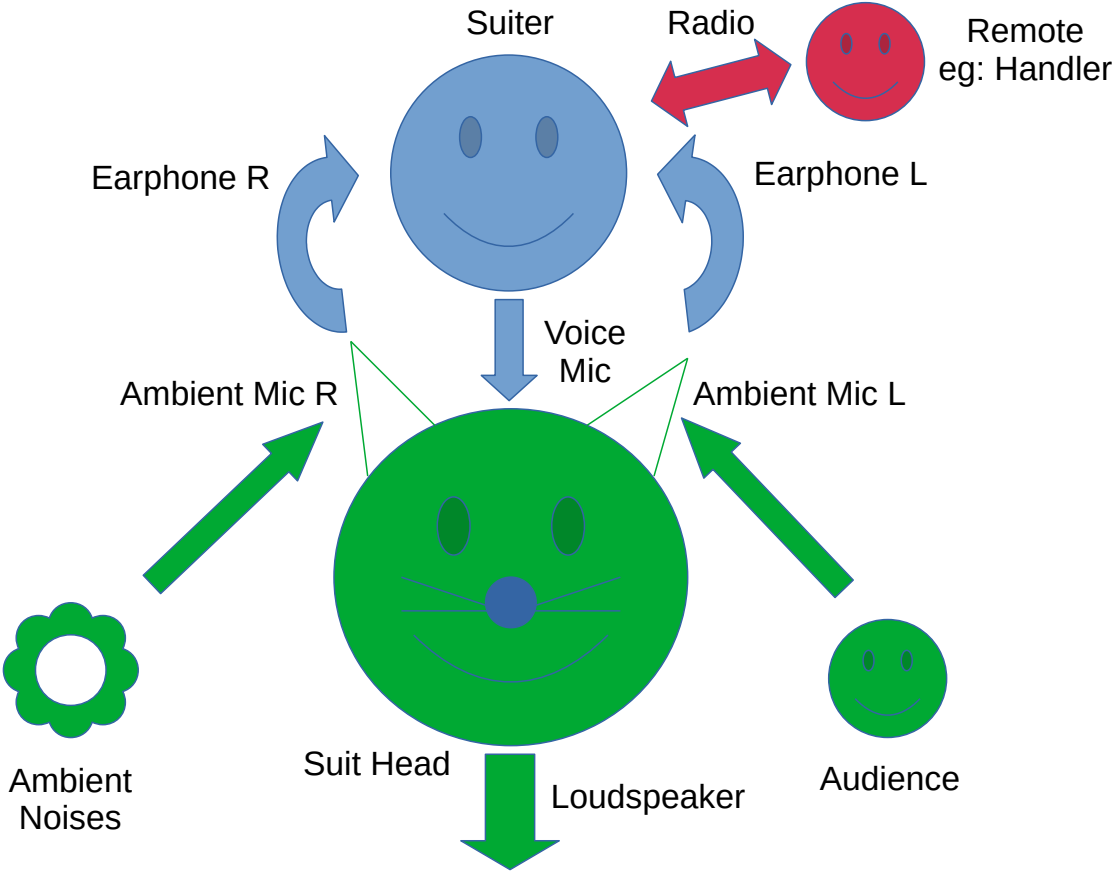
- **Demonstration**

The Concept

Core Idea:

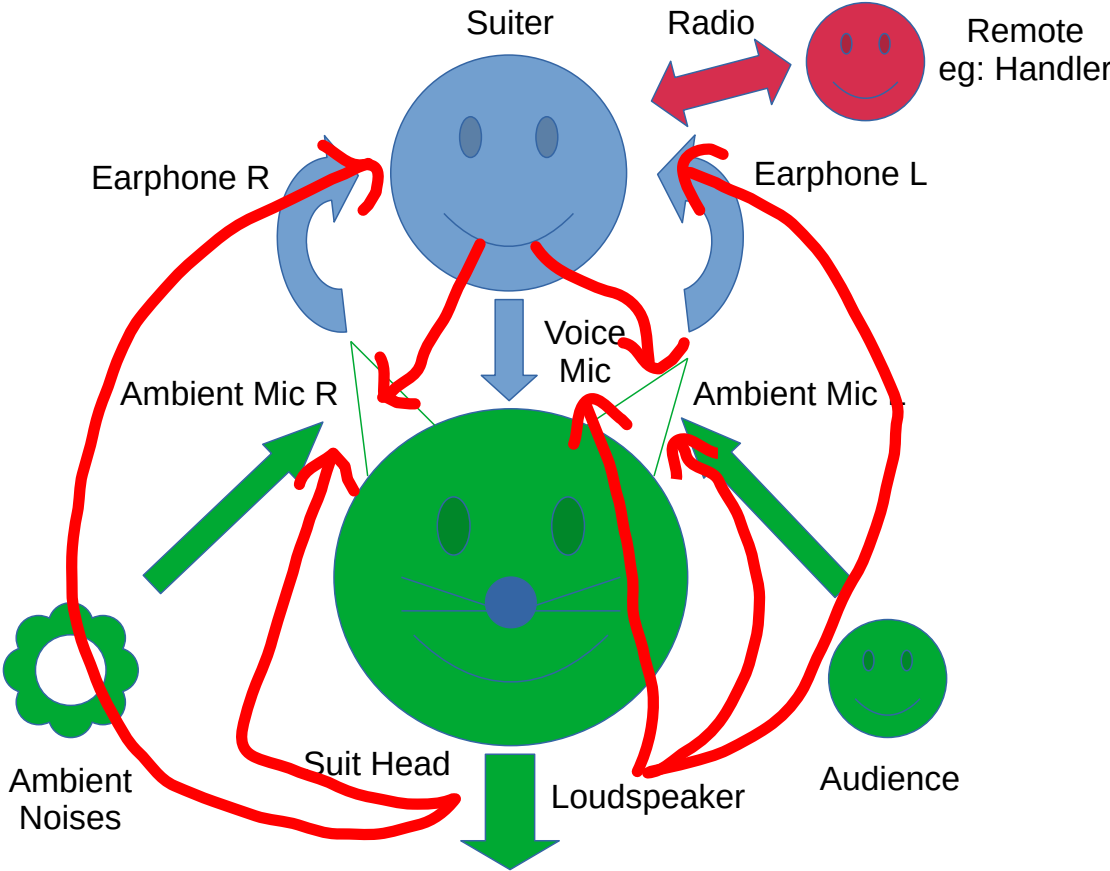- **Make the suit perceptually transparent to the suiter**

Core Functions:

- **Speak clearly**

- **Hear clearly**

- **Communicate (~invisibly) with remote handler + others**

What's With all those Red Squiggly Lines ?

Those red, squiggly lines are BIG TROUBLE!

- **They represent undesired sound paths**
    - Much of my R&D has been about methods to reduce their impact
- **Why they are troublesome is explored in the next section**

Content

- **The Goal**

- **The Concept**

- **Why is this so Hard?**

  - Interference Effect

  - Larsen Effect

  - Close-Talking, Directional Microphone Issues

  - Suit Head and Loudspeaker Resonances

  - Head-Related Transfer Functions

  - Putting it all Together

- **Wrapping-Up**

- **Demonstration**

Why is this so Hard ?

It's just a couple of microphones and a loudspeaker, isn't it?

→ The fact that you can't buy such a device should be a hint...

It boils down to two things:

- The human speech production system can be **interfered** with, making it hard/impossible to speak

- Sound systems **howl** when the microphone is put too close to the loudspeaker

And that is exactly what happens with voice re-inforcement in-suit (Yay!)

We'll now look into these challenges in detail

Content

- **The Goal**
- **The Concept**
- **Why is this so Hard?**
  - Interference Effect
  - Larsen Effect
  - Close-Talking, Directional Microphone Issues
  - Suit Head and Loudspeaker Resonances
  - Head-Related Transfer Functions
  - Putting it all Together
- **Wrapping-Up**
- **Demonstration**

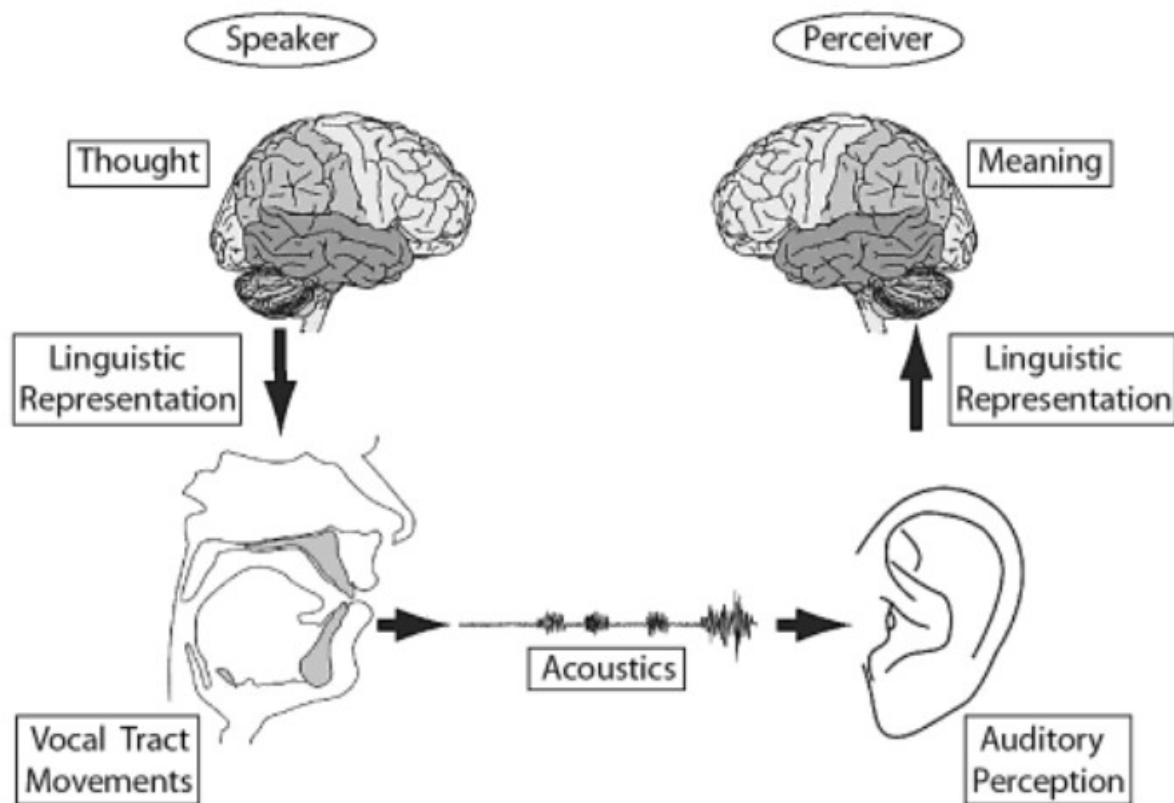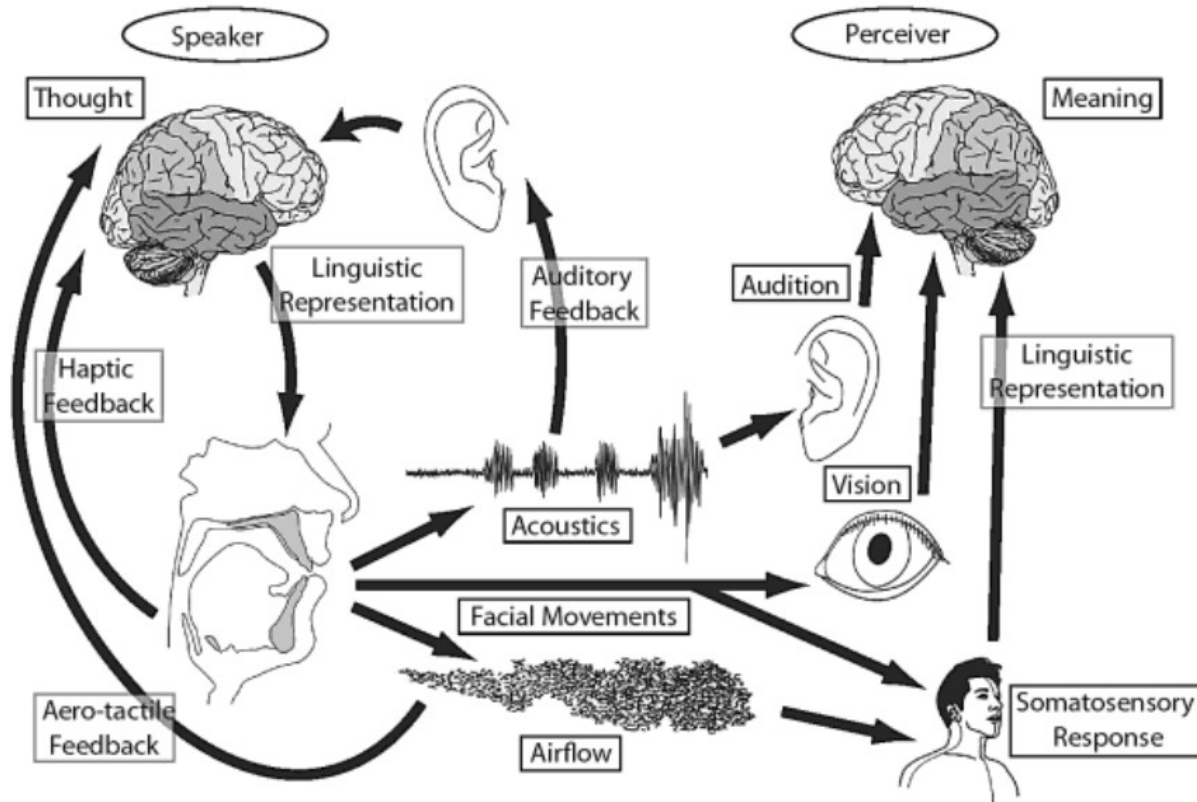Figure 1.1, Articulatory Phonetics, Gick, Wilson and Derrick

Figure 1.2, Articulatory Phonetics, Gick, Wilson and Derrick
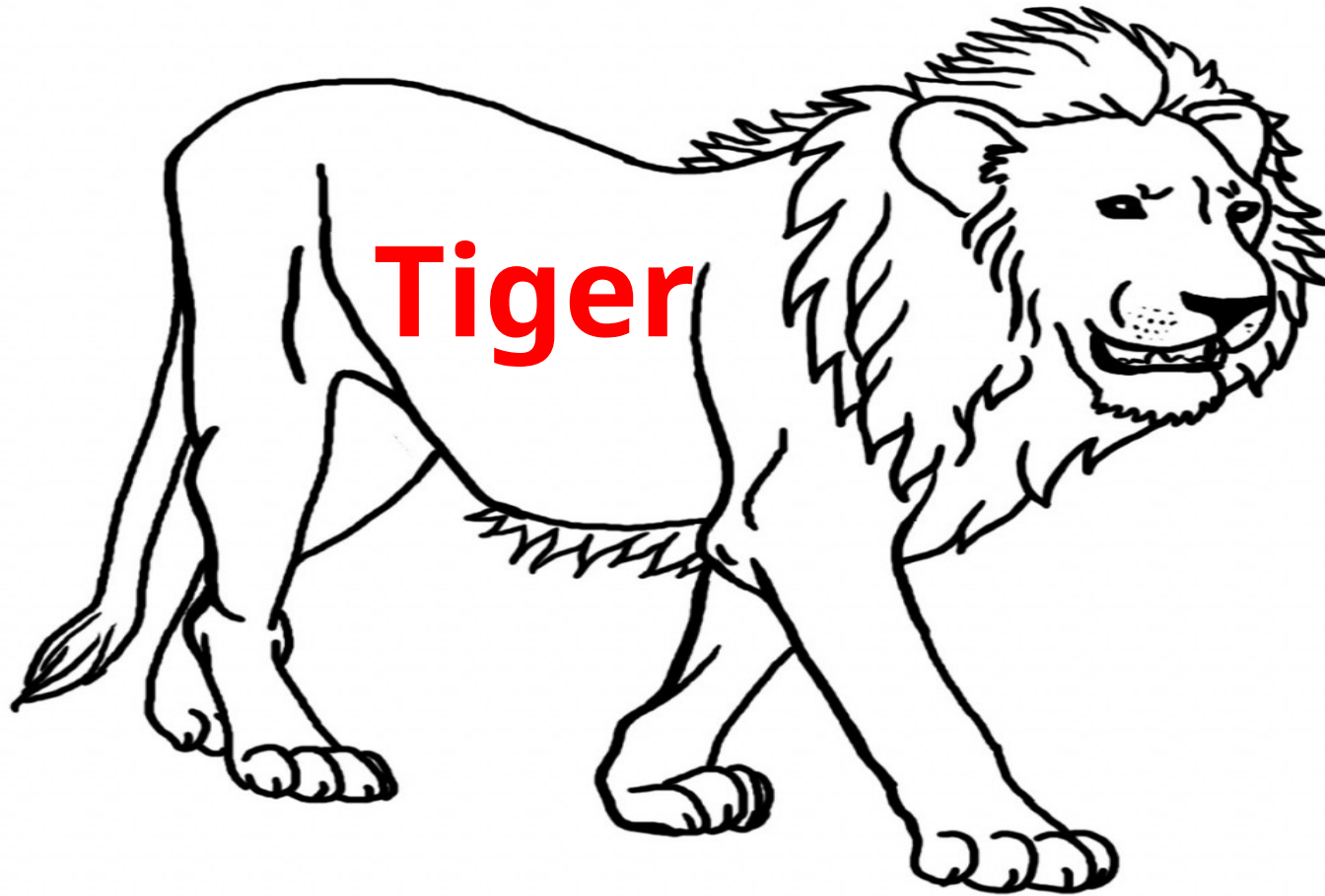
Interference: Hacking the Speech Chain

If something is complex, it can be messed with!

- **Messing with any of the feed-back loops should do it**
- **These loops are multi-modal!**

Let's have some fun with this...

Tiger

RED

McGurk Effect

McGurk effect - Auditory Illusion - BBC Horizon Clip
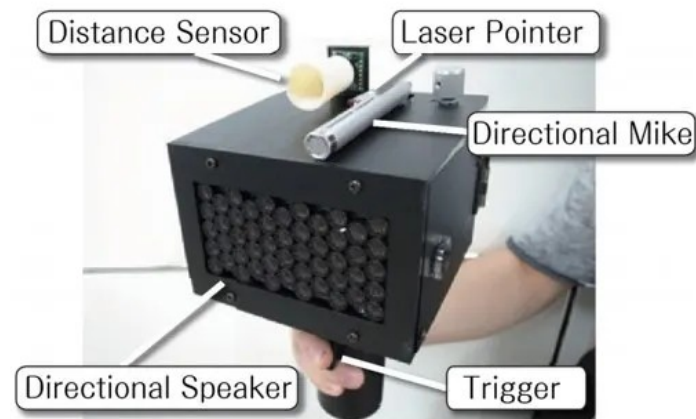https://www.youtube.com/watch?v=2k8fHR9jKVM

The Speaker can also be Messed With !

Delayed Auditory Feed-back - DAF (also know as "delayed side-tone")

- **Messes with the auditory feed-back path → DEMO!**

- **Usually noticeable between 50-200 ms**

  - Young children, the elderly and possibly the autistic can tolerate more delay

- **Causes mental stress**

- **Can cause serious issues with speaking**

  - But can help some forms of stuttering and cluttering

  - Has been weaponized: SpeechJammer

    https://en.wikipedia.org/wiki/Delayed_Auditory_Feedback
    https://www.smithsonianmag.com/science-nature/the-sound-gun-that-will-leave-you-speechless-115378328/

Delayed Auditory Interference

This is the bugbear we're dealing with in costume voice:

- **You will hear yourself back via the loudspeaker with delay**

- **Some delay is inherent in digital signal processing**

- **Most of the interesting voice effects add too much signal delay**

  - More (realistic) computation won't solve this for many algorithms!

The question is: can we break the auditory feed-back loop?

→ Why, yes! But it ain't a walk in the park
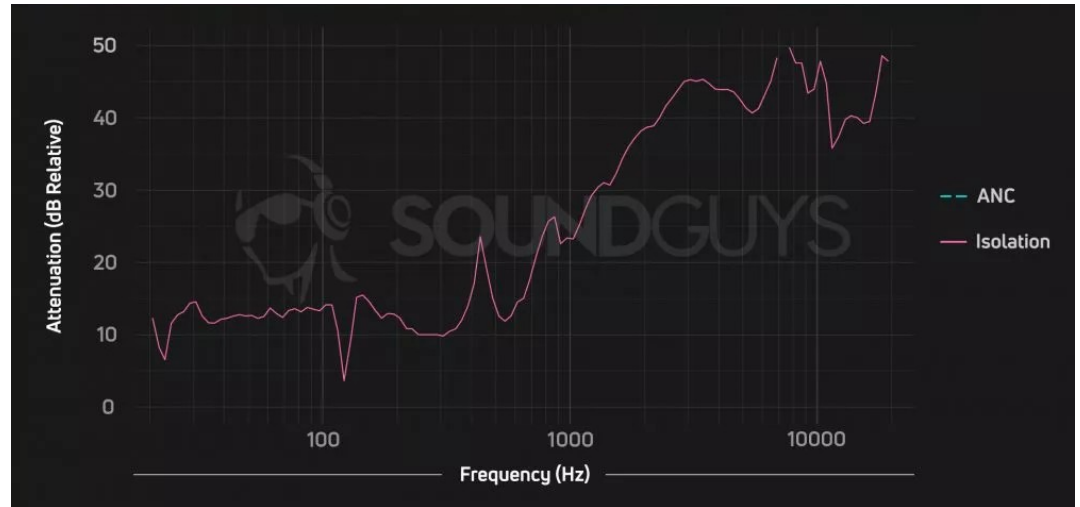
Sound Paths – More Complete – More Trouble
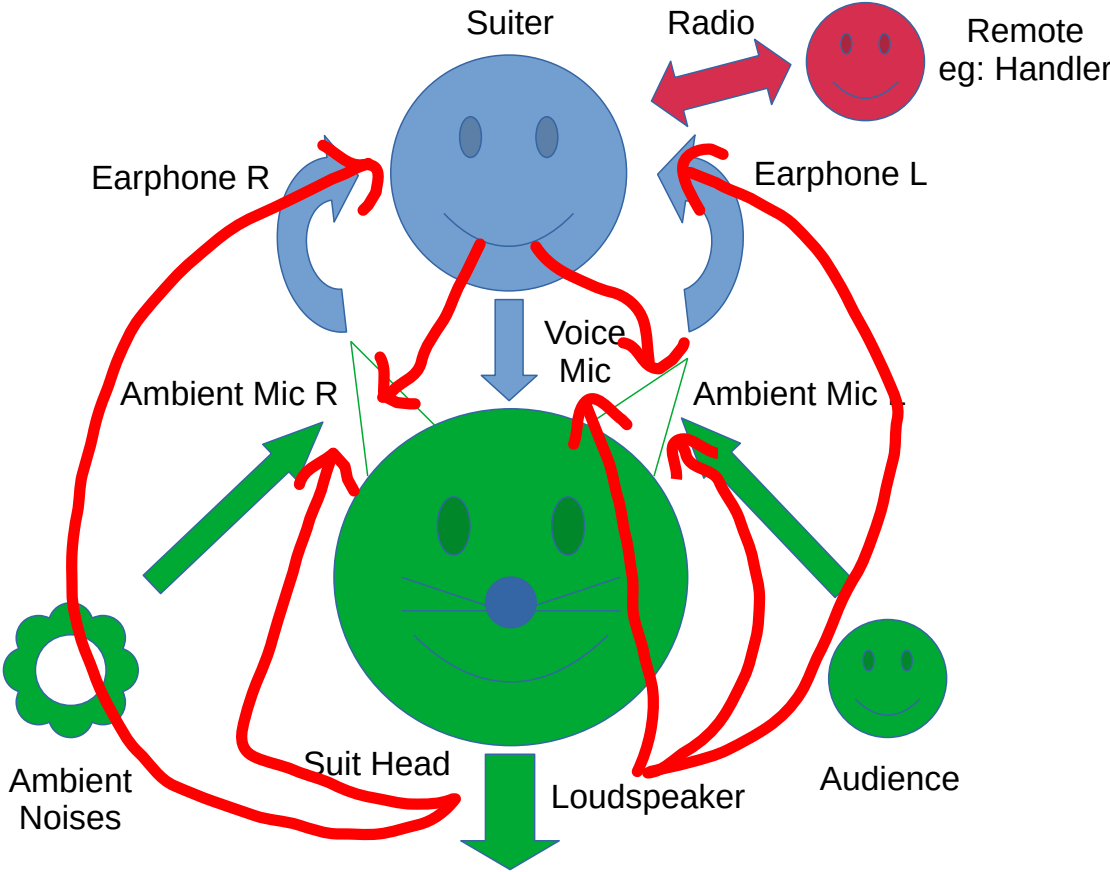
Loudspeaker to Ears Path

Easy enough to resolve:

- **In-Ear Monitors (IEM)**
  - Aka: drummer earphones
- **Offer a good amount of passive sound isolation**
- **Stay put quite well**
- **Quite flush to the head**
- **Readily available (€100)**
- **(Hope you like ear wax...)**

SoundGuys – Shure SE215 Review
https://www.soundguys.com/shure-se215-review-77980/

Loudspeaker to Ambient Microphones Path

This one is really hard to tackle…

- **Delay causes interference**

- **The loudspeaker is loud**

  - Who saw that one coming?

  - We can't turn the ambient mics down, we still need to hear the ambient!

Requires the use of high-performance, adaptive filters

- **Adaptive Echo Cancellers (AEC), but in HARD mode**

  - Operates in near-continious double-talk

  - Near-end and far-end highly correlated

- **We'll explore what this all means next…**

AEC Explained

The goal is to model the acoustic path "F" between the loudspeaker and microphone → This is what every telephone exchange, cell phone, hands-free phone, and teleconference software has to do (and you've heard it go wrong often enough...)

- **Using the estimated $\hat{F}$, we know (approximately) what the loudspeaker signal is at the microphone**

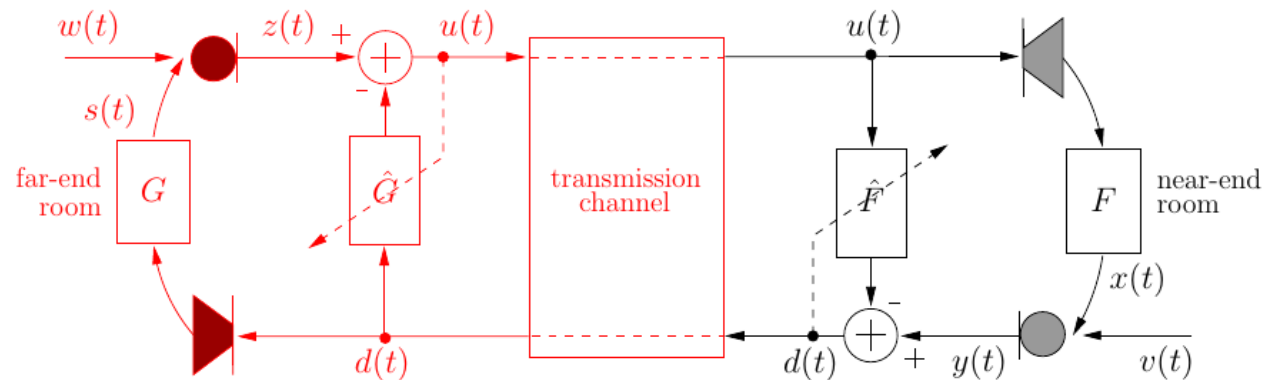- **This allows us to substract the loudspeaker contribution from the microphone signal**

Figure 1.1 - PhD Thesis - Toon van Waterschoot

So, does this actually work? → **DEMO**

Yes, quite nicely even… But it still **fails badly?!**

- The filter path estimation uses various forms of gradient descent algorithms. These make assumptions about the nature of the signals:
  → They must be both **white** and **uncorrelated**

- In our case: the signals are speech (**anything but white**) and the interference (near-side speech) is **highly-correlated** AND under continuous **double-talk**
  → more like AFC, which is much harder!

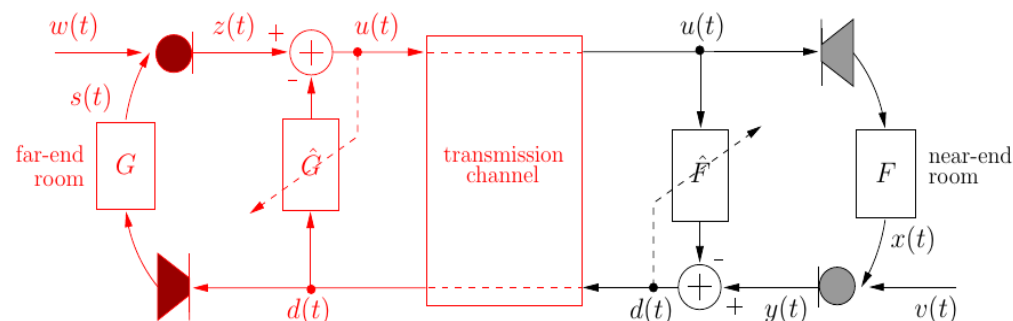- The problem is that the adaptation gets **confused** about what the path F really is

Figure 1.1 - PhD Thesis - Toon van Waterschoot

This can be resolved using **AFC**-like techniques, discussed next, using:

- Auxiliary filters to make the loudspeaker signal white (which will also **decorrelate** it) without disturbing the loudspeaker signal

- Make the adaptation **robust** to double-talk (done in the demo we'll show later)

Content

- **The Goal**

- **The Concept**

- **Why is this so Hard?**
  - Interference Effect
  - Larsen Effect
  - Close-Talking, Directional Microphone Issues
  - Suit Head and Loudspeaker Resonances
  - Head-Related Transfer Functions
  - Putting it all Together

- **Wrapping-Up**

- **Demonstration**

On the Topic of Feed-Back…

- **Larsen effect is another challenge**
  - The voice microphone and loudspeaker are very close together in a suit head
  - You just know it's going to screech like a wounded banchee on speed…

- **Again an unwanted, positive feed-back loop at work**
  - Yes, we can do something about this
    No, it is not simple

Feedback loop between input and output

AMP

www.rahulsamuel.com
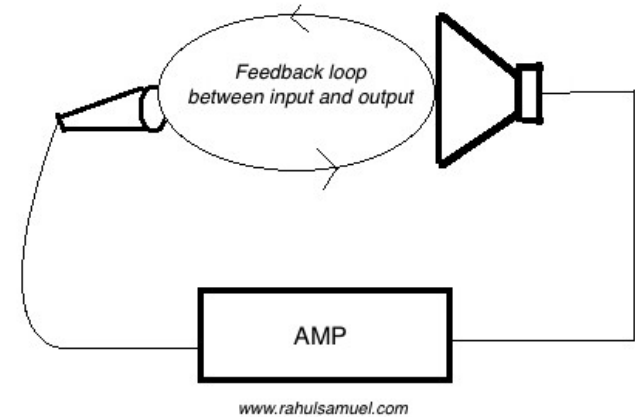
AFC Explained

Acoustic Feed-back Cancellation is conceptually similar and simpler vs AEC:

- **There is no far-end**

- **Estimate F to remove the loudspeaker contribution from the microphone signal**

If we can reduce the loudspeaker contribution, we can reduce howling and also allow **more gain** (play louder)

Great! But, double-talk and correlated signals are definitely the Big Issue here…
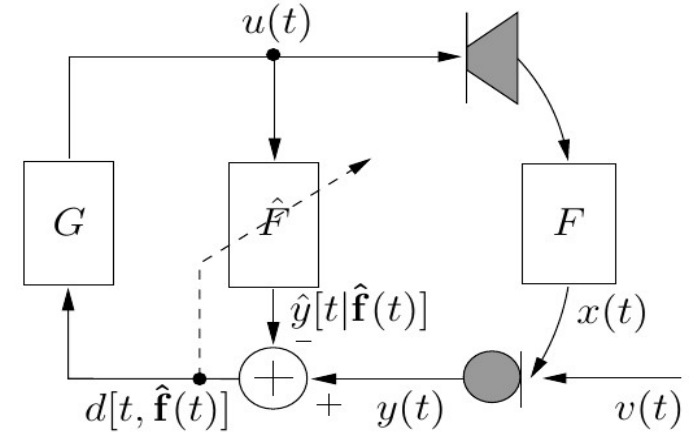
→ The simple approach fails in practice

Figure 2.11 - PhD Thesis - Toon van Waterschoot

We need to whiten (decorrelate) the signals to make this work

- **Signals are not stationary**

- **Should not deteriorate the loudspeaker signal (perceptively)**

Needs (near) state-of-the-art algorithms to do this well!
→ Such as PEM-AFROW
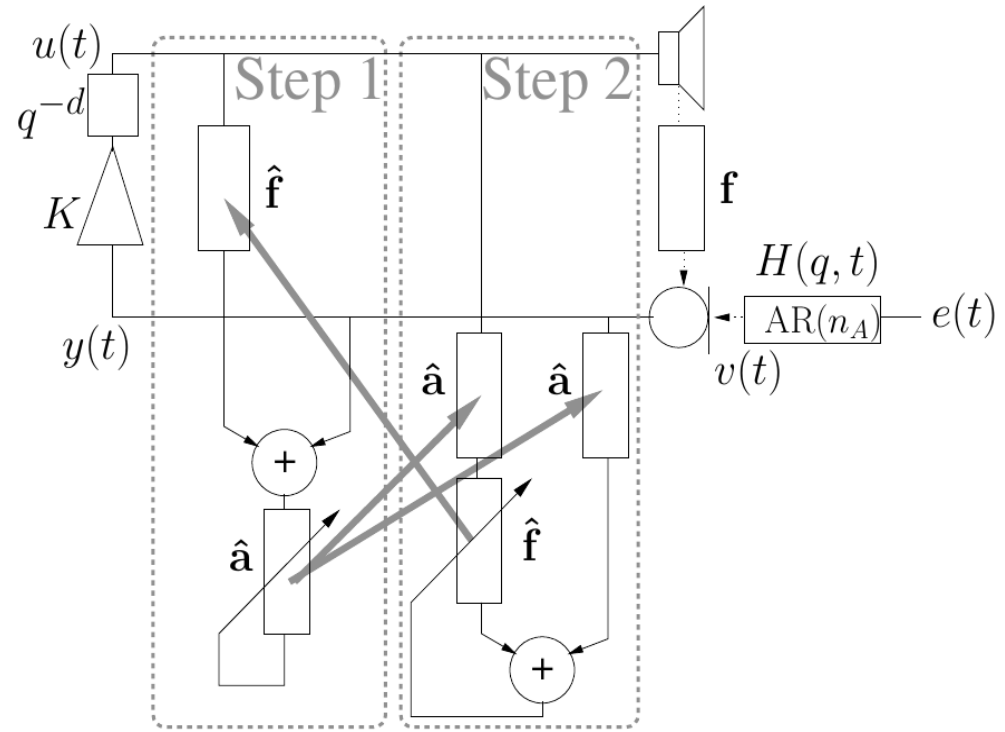

Figure 9.5 – PEM-AFROW Adaptation Stages – PhD Thesis – Toon van Waterschoot

Content

- **The Goal**
- **The Concept**
- **Why is this so Hard?**
  - Interference Effect
  - Larsen Effect
  - Close-Talking, Directional Microphone Issues
  - Suit Head and Loudspeaker Resonances
  - Head-Related Transfer Functions
  - Putting it all Together
- **Wrapping-Up**
- **Demonstration**

Need for a Good, Directional Microphone

- **Need to reduce loudspeaker feed-back, as well as background noise**
  - No space for a shotgun mic
- **Closer to the sourse = better SNR!**
  - Move close to the lips!
  - Use a directional microphone
- **In practice:**
  - Cardioid, head-worn microphone
    Though: usually figure-eight
  - Close to the lips = within 2 cm or so

RØDE NTG-2 Shotgun Microphone

AKG CM311 Head-Worn Microphone
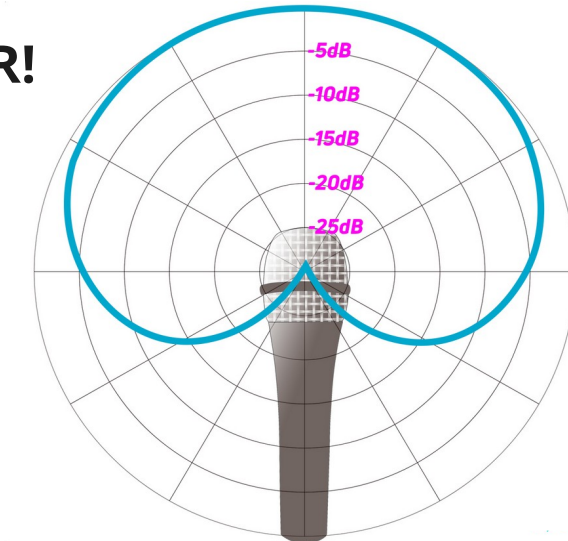
Cardioid - SoundGuys
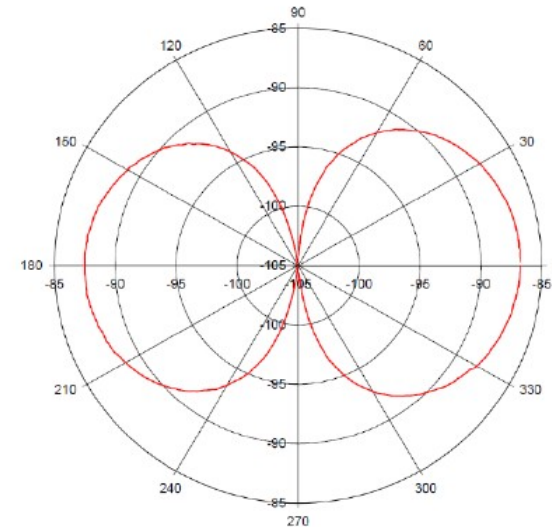
Figure Eight - Knowles Acoustics

Close-Talking, Directional Microphones

Shure Beta 57A

- **Works very well at reducing background noise! → DEMO**

- **But, colors the sound something fierce: → DEMO**

  - Booming (cardioid and figure-8 are pressure gradient microphones)

  - Plosives, sibilance + breathing

  - Unsatisfactory high-frequency behaviour

  - Sounds different than a microphone at a little distance (oral-nasal mix + radiation patterns)

Much can be fixed:

- **High-pass filter to reduce booming**

- **EQ can add high-frequency peaking (around 6 kHz) + HF roll-off to sound more natural for voice**

- **Pop filter / wind screen + de-esser (sibilance control) is definitely needed!**

An omni-directional mic will be better for this aspect, but still needs work
→ And we loose the directivity!
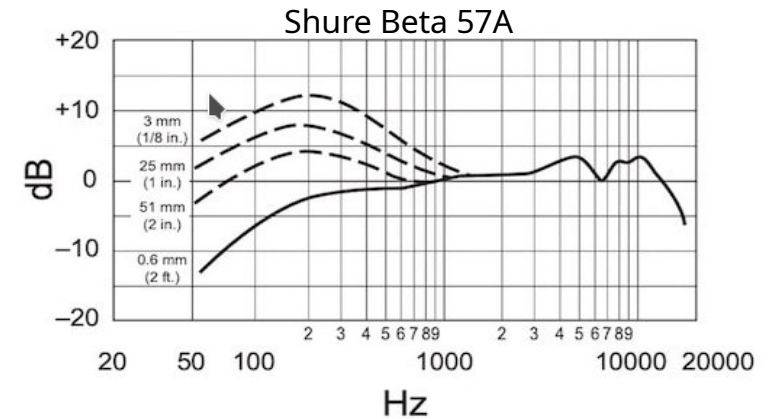
What our Parametric EQ Looks Like !
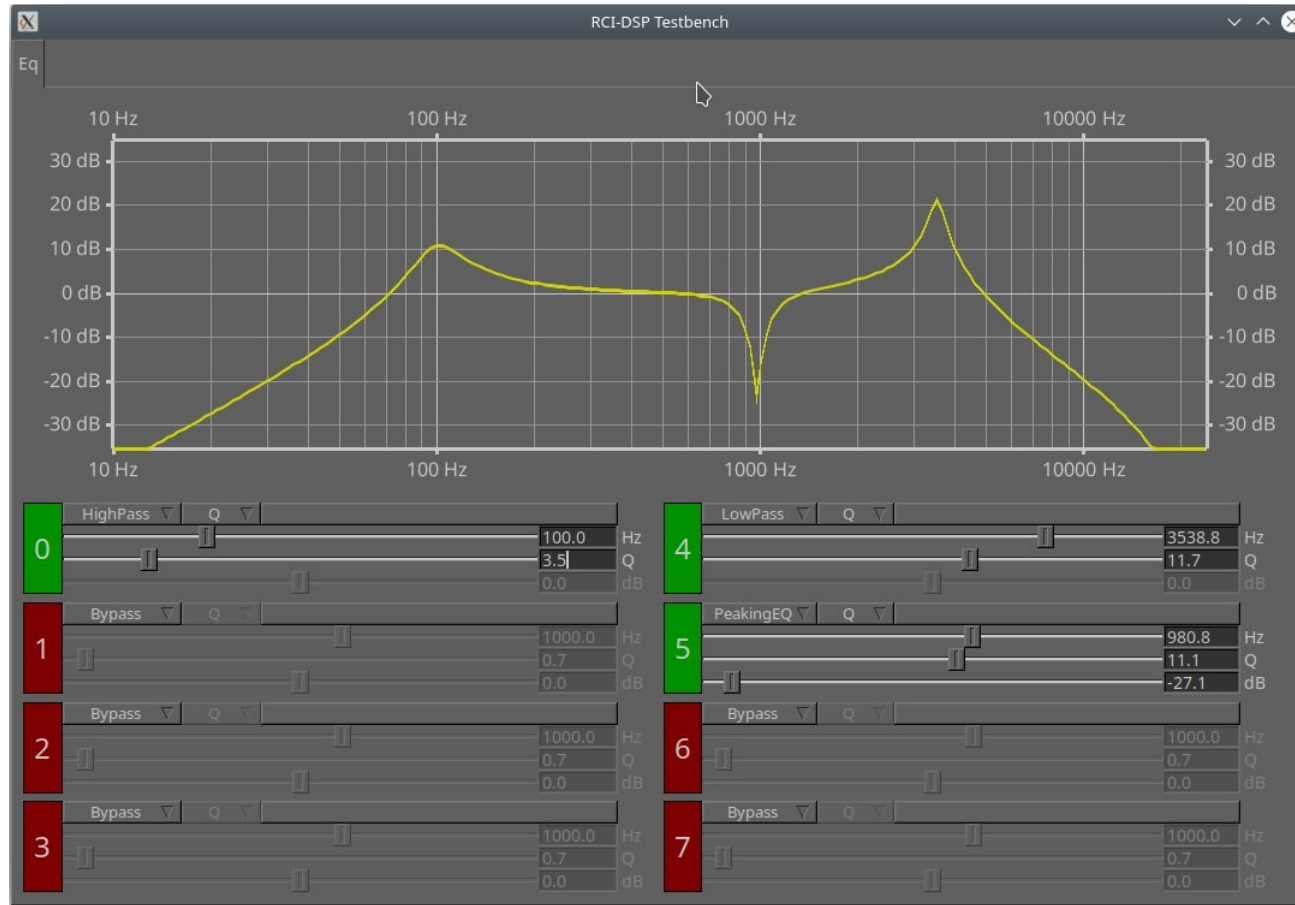
Content

- **The Goal**
- **The Concept**
- **Why is this so Hard?**
  - Interference Effect
  - Larsen Effect
  - Close-Talking, Directional Microphone Issues
  - Suit Head and Loudspeaker Resonances
  - Head-Related Transfer Functions
  - Putting it all Together
- **Wrapping-Up**
- **Demonstration**

(Anti-)Resonances

(Anti-)resonances can occur in:

- **Loudspeaker enclosure**
  - Poor tuning/damping
  - Enclosure vibrates
  - Driver ill-matched
  - Poor-quality driver
- **Head base**
  - A hard material may? support audible resonances

These can sound bad when your voice hits them!

- **Can't always fix at the source**
- **Some EQ will go a long way!**
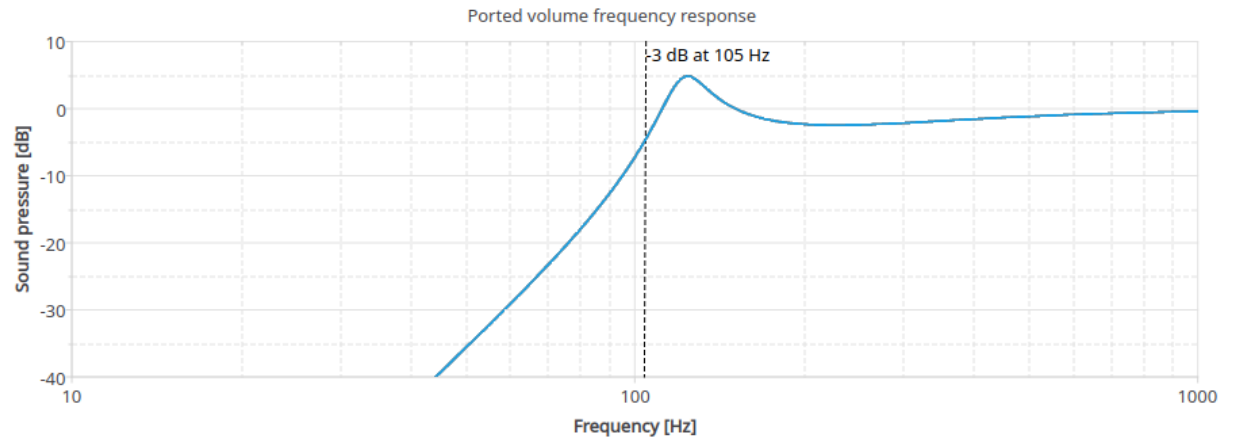
Ported volume frequency response

-3 dB at 105 Hz

Content

- **The Goal**

- **The Concept**

- **Why is this so Hard?**
  - Interference Effect
  - Larsen Effect
  - Close-Talking, Directional Microphone Issues
  - Suit Head and Loudspeaker Resonances
  - Head-Related Transfer Functions
  - Putting it all Together

- **Wrapping-Up**

- **Demonstration**

Head-Related Transfer Functions

So, with good earphones and decent, directional microphones for hearing the ambient sound, it's going to sound good, right?

You guessed it: nope

- Our ears are complex filters that help us localize sounds in 3D

- With earphones, we by-pass all of this filtering (that we are used to since birth) and it sounds... unpleasant

- (Also: rubbing noises when petting, but that is another story)
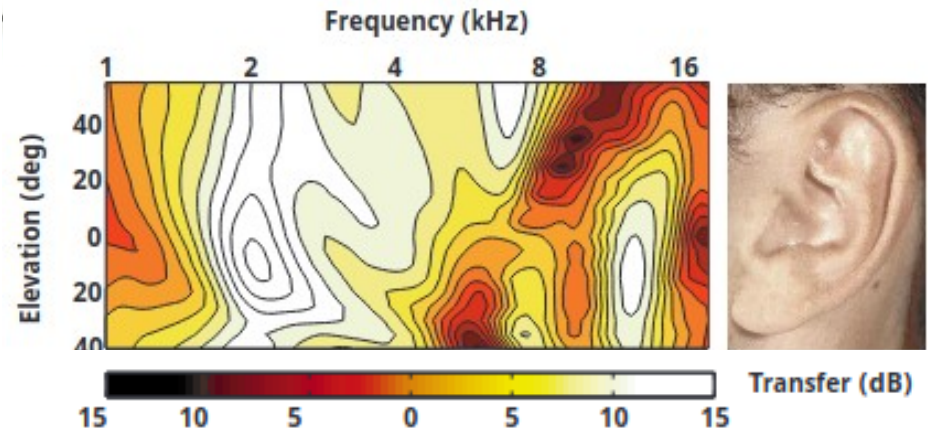
Size and shape of: the head, outer ears, ear canals...
→ affects these filters for each individual

- **Generally:**

  - A boost in the 2-5 kHz range

  - Primary resonance of +17 dB at 2.7kHz

- **Filter depends on sound elevation as well**

  - Helps us locate sound in the up-down direction

t.bone LC-97 TWS
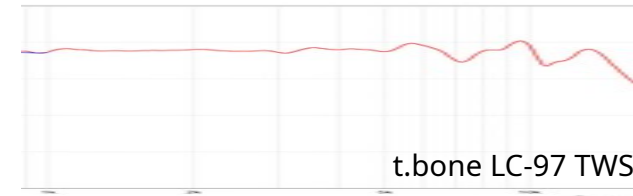
"Relearning Sound Localization with New Ears," Hofman et. al.

Head-Related Transfer Functions

What to do about this?

t.bone LC-97 TWS

- **If we pick one elevation, we can make it sound pleasant enough But we we loose most elevation information**

  - Some experimenting shows this to be fine in many/most situations

  - We still get azimuth (left-right, front-back) information

- **The harshness can be solved with a little filtering (EQ)**

  - Emulate the bump at 2.7 kHz

  - Add a general, gentle LP roll-off

    - This emulates a ribbon microphone response, which often sound pleasant for voice work
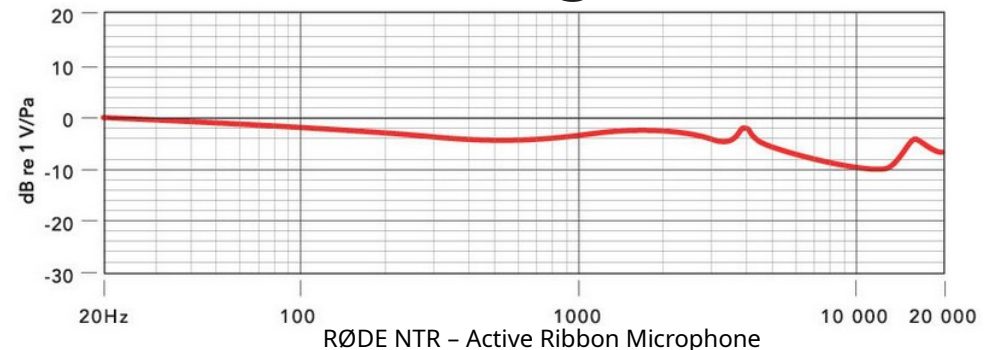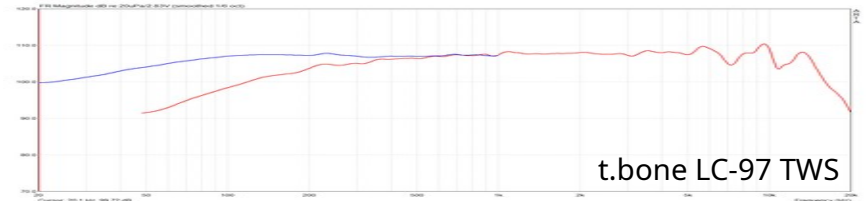
RØDE NTR – Active Ribbon Microphone

Content

- **The Goal**
- **The Concept**
- **Why is this so Hard?**
  - Interference Effect
  - Larsen Effect
  - Close-Talking, Directional Microphone Issues
  - Suit Head and Loudspeaker Resonances
  - Head-Related Transfer Functions
  - Putting it all Together
- **Wrapping-Up**
- **Demonstration**

Putting it all Together: Audio Flow Diagram
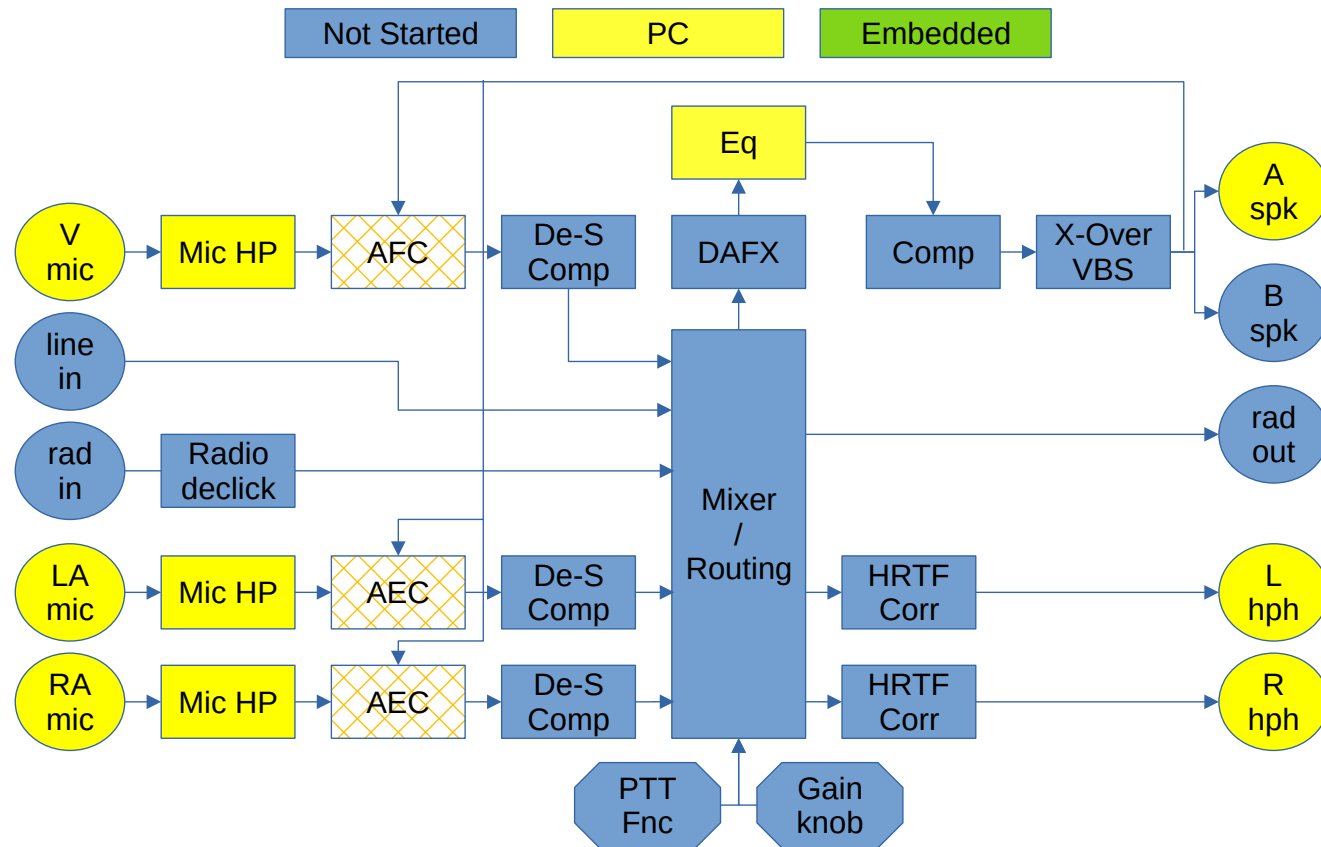
Content

- **The Goal**
- **The Concept**
- **Why is this so Hard?**
- **Wrapping-Up**
- **Demonstration**

Hardware

- **First electronics module prototype is ready**
  - No change since last year

- **First suit integration prototypes are ready**
  - AlphaZero R&D head base
  - Audio integration test rig

→ **We're in pretty good shape!**

Where is the Project at ?

Software

- **DSP algorithm R&D**
  - 60% done (depends...)
- **Embedded platform code**
  - 5% done (basic RTIC + USB comms)
- **Communication protocol**
  - 70% done (already runs embedded!)
- **User control program (PC/Mac)**
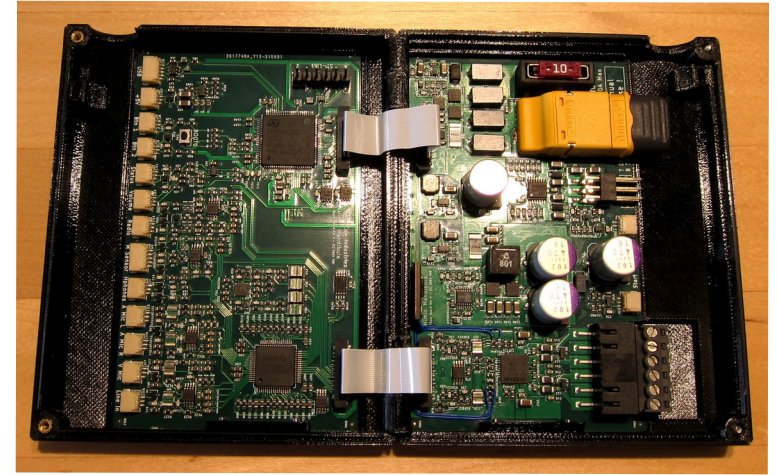 - 30% done

```
Language               Files    Lines   Blanks  Comments    Code Complexity Complexity/Lines

Rust                      51    14121     1311      5816    6994        565           334.84
C                         15     2615      349       284    1982        284           205.28
C Header                   8      566       97       197     272          4             4.60
CMake                      4      486       83       102     301         40            42.40
Makefile                   3      612      107       156     349         17            11.63
License                    2       41       10         0      31          0             0.00
C++ Header                 2      665       90        92     483         51            21.04
gitignore                  2       54        0         0      54          0             0.00
Autoconf                   2       92       11        26      55          9            19.15
Markdown                   1      245       71         0     174          0             0.00
YAML                       1       27        6         0      21          0             0.00
Shell                      1      275       40        34     201         31            15.42
Plain Text                 1       32        3         0      29          0             0.00
JSON                       1       27        0         0      27          0             0.00
TOML                       1       32        5         4      23          0             0.00
Python                     1      143       25         6     112         26            23.21
C++                        1       80        9         7      64          6             9.38

Total                     97    20113     2217      6724   11172       1033           686.95

Estimated Cost to Develop (organic) $340,512
Estimated Schedule Effort (organic) 9.13 months
Estimated People Required (organic) 3.31
Processed 791618 bytes, 0.792 megabytes (SI)

tomr@Kingfisher:~/workspace_RoboCow_Rust/rci-dsp$ cd ../rci-dsp-test/
tomr@Kingfisher:~/workspace_RoboCow_Rust/rci-dsp-test$ ./sloc.sh
Language               Files    Lines   Blanks  Comments    Code Complexity Complexity/Lines

Rust                      54     7279      763      1786    4730        118            94.72
Protocol Buffers          12      405       65       123     217          0             0.00
TOML                       1       58       11        25      22          0             0.00
JSON                       1       45        0         0      45          0             0.00
gitignore                  1        3        0         0       3          0             0.00

Total                     69     7790      839      1934    5017        118            94.72

Estimated Cost to Develop (organic) $146,913
Estimated Schedule Effort (organic) 6.64 months
Estimated People Required (organic) 1.97
Processed 348104 bytes, 0.348 megabytes (SI)
```

What do We Hope to Get Done during 2023 ?

In one word: Software

- **Complete enough of the DSP algorithms for a key decision gate**
  - Pass/Fail decision for the whole project: does this work well enough, on reasonable hardware?
- **Embedded platform**
  - Develop the core functionality on the embedded platform
    - Communications, device drivers, DMA-based audio data flow, …
  - Port the DSP algorithms (and add plaftform-specific optimizations for FFT, FIR, etc…)
- **User control program**
  - Re-factor the whole thing to work with the embedded target
  - Add missing control tabs and core features

Goal is to demonstrate a sufficiently complete, working system (even if still running on a PC)

- **Also to potential, professional clients**
  → big part of financial viability (also a key decision gate)
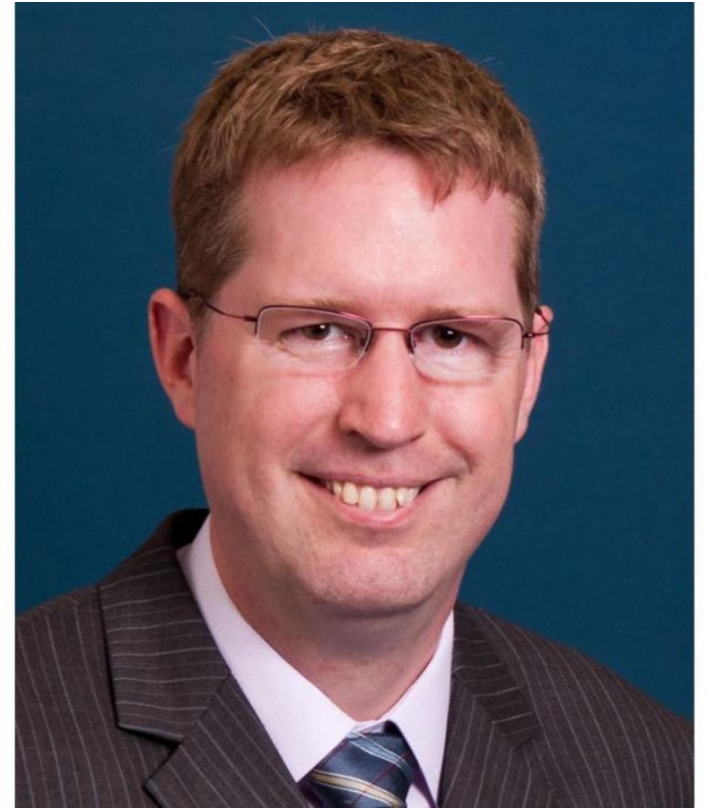
We ? Surely you Jest !

No! To my sheer and utter amazement, and don't ask me why or how, RoboCow Industries has a loose collection of **volunteers** from **all over the world** supporting our efforts!

We ? Surely you Jest !

Dr. Donald Derrick – **New Zealand**
(A normie friend, but I'm sure he'd resent that label and be the first to prance around in a fursuit given half the chance... Just to see what it's like)

- **My Virgil into the darkest of the Humanities**
  - Speech science, linguistics, anthropology, statistics...
  - We've done science and invented stuff together, down-under and in the Kiwi's nest

- **Helps me make sense of "humans" and their shockingly shoddy manufacturing tolerances**

- **We go back a long, long time, across continents...**

We ? Surely you Jest !

NBF **(Normie Business Friend) – Belgium**

- **Philosophy**
- **Business advice**
- **Sanity checks & keeping my feet on the ground**
- **Generously gave me a week of his time !**
  (He's the CTO of a good-sized SME)
- **I occasionally blow-up his designs in the lab…**
  That's how we met and got talking ☺

He actually attended this panel last year, and…

- **Saw the technicolour line-up for the photoshoot**
- **Got hugged by a giant, plush, pink otter**
- **We're still friends after all of that!**

IMAGE NOT FOUND
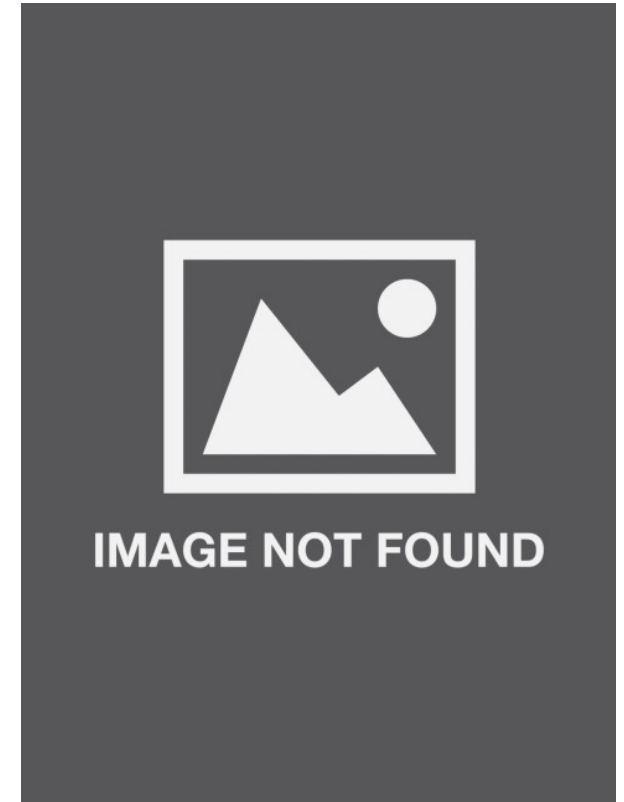
We ? Surely you Jest !

Tisha Tiger – **France**

- **"Manages" the insane menagerie at Highway to Tail https://highwaytotail.com/**

- **Kindly includes my ramblings in his excellent videos**
  - Go and watch them! Many have English sub-titles (CC button)

- **Believe it or not, this is what has helped give RoboCow credibility in the real world! (And now the family knows what a loon I am...)**

We ? Surely you Jest !

Juravenator **– Belgium**

- **Offered to help-out last Flüüfff...**

- **Oh, WOW! Where to even start!**

  - Re-wrote the entire communication stack with ProtoBuffers, on embedded (no_std + no_alloc), amongst other feats of Rust

  - Always there to help out when I get stuck (again)

  - Clear opinions and constructive feed-back (← rare stuff, that!)

  - Converts coffee into sarcasm with truly remarkable efficiency

- **Refers to the talking cow as his "sanity project" Consider that for a moment...**

  - Hint: Jura works in big corp software development / IT
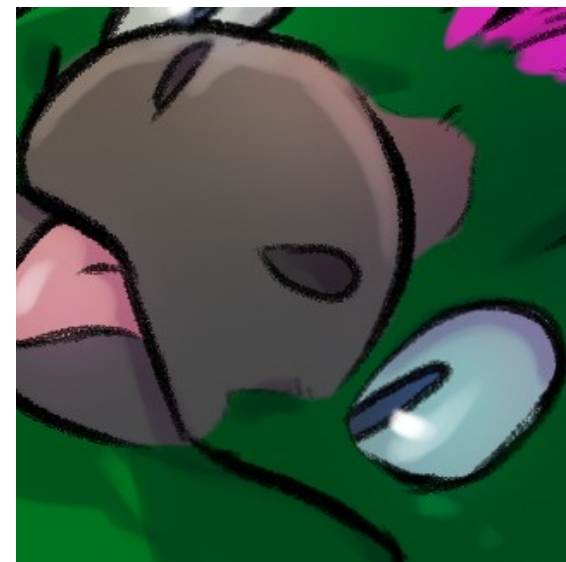
We ? Surely you Jest !

Over the years, we've gotten help from many others as well! Notable instances:

- **Rua – The Netherlands**
  - Convinced me to use Rust for all of this, and helped-out when I imediately got stuck in lifetime hell
- **Ferix – The Netherlands**
  - Part of the Rust Conversion Tag Team, looked at my code and shared code to help me get started
- **Folks in the Rustacean Furs Telegram and Rust Matrix channels**
  - Helped with issues when I got started
- **Sibs – Germany**
  - Helped clarify the realities and economics of fursuit making to NBF and myself

Plus: all of your questions, feed-back and discussions after the panel
→ It all goes into the development process!

Thank YOU

Content

- **The Goal**

- **The Concept**

- **Why is this so Hard?**

- **Wrapping-Up**

- **Demonstration**

Demonstration

**Eh, well, sure... If you insist...**

- **What we have here is a research step "frozen in time"**

- **IT IS NOT DONE!**

  - AEC algorithm performance is ... not so good

  - AFC algorithm performance is ... well, it does *something*

  - None of the "pleasantness" filters are in place yet

  - Sibilance control and compressors are still missing

  - Latency feels like jet lag (mainly due to PC latency in R&D platform)

- **With that said... Let's embarras ourselves and have some fun!**

Demonstration

A volunteer is needed!

- **Brave, strong, cute**

- **Experienced suiter**

- **Some affinity with technology**

- **Not afraid (see brave) of my R&D stuff, as there are some RISKS!**

- **You will have to stick oversized, silicone objects in your ears!**
  → people will see your earwax…

STAND BACK
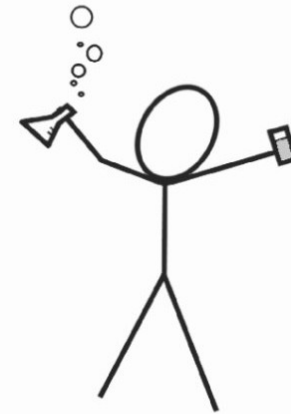
I'M GOING TO TRY SCIENCE

ROBOCOW Industries