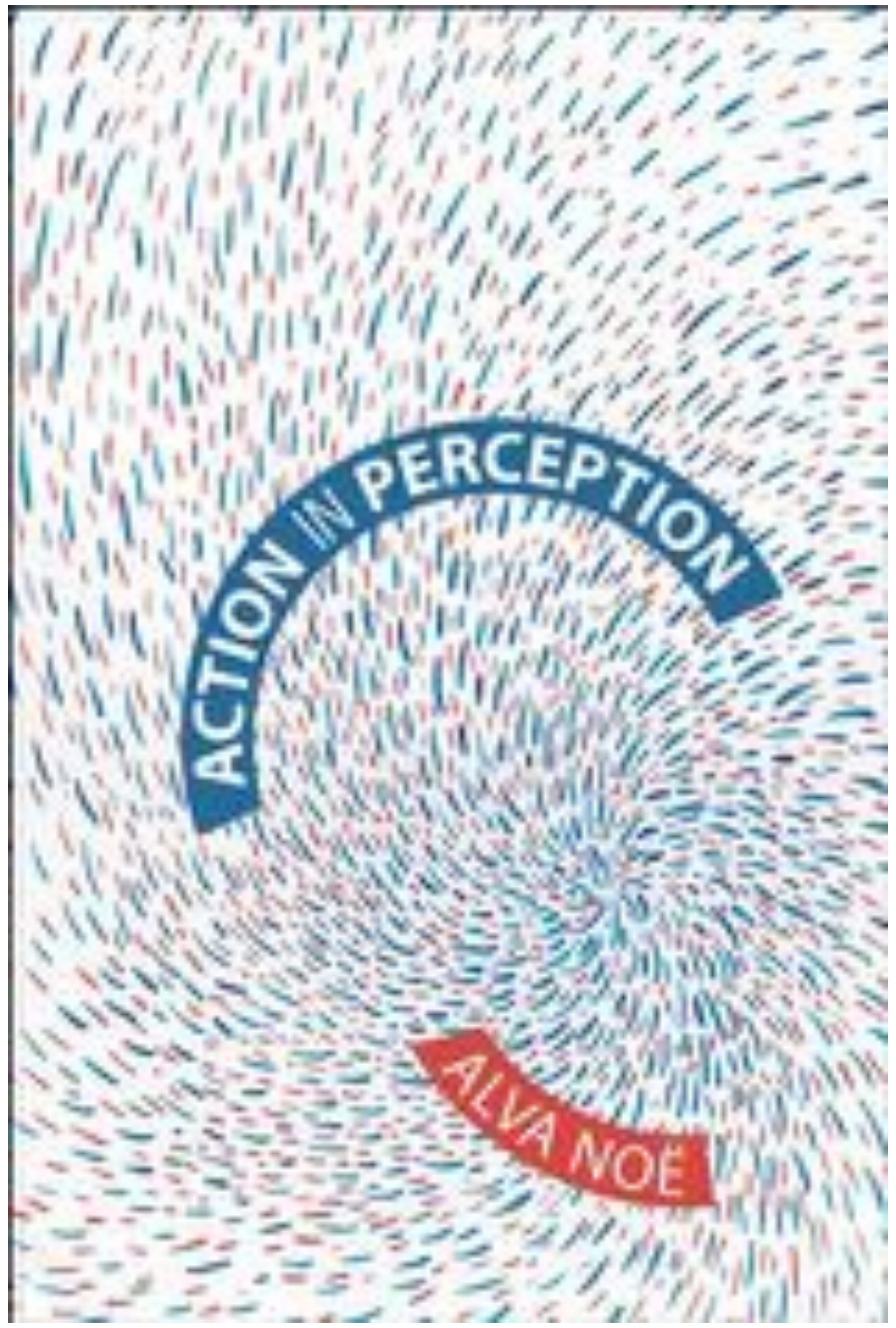
Aalto University
School of Arts, Design
and Architecture

action, perception and sonic interaction design

AHNE - audio haptic navigation environment

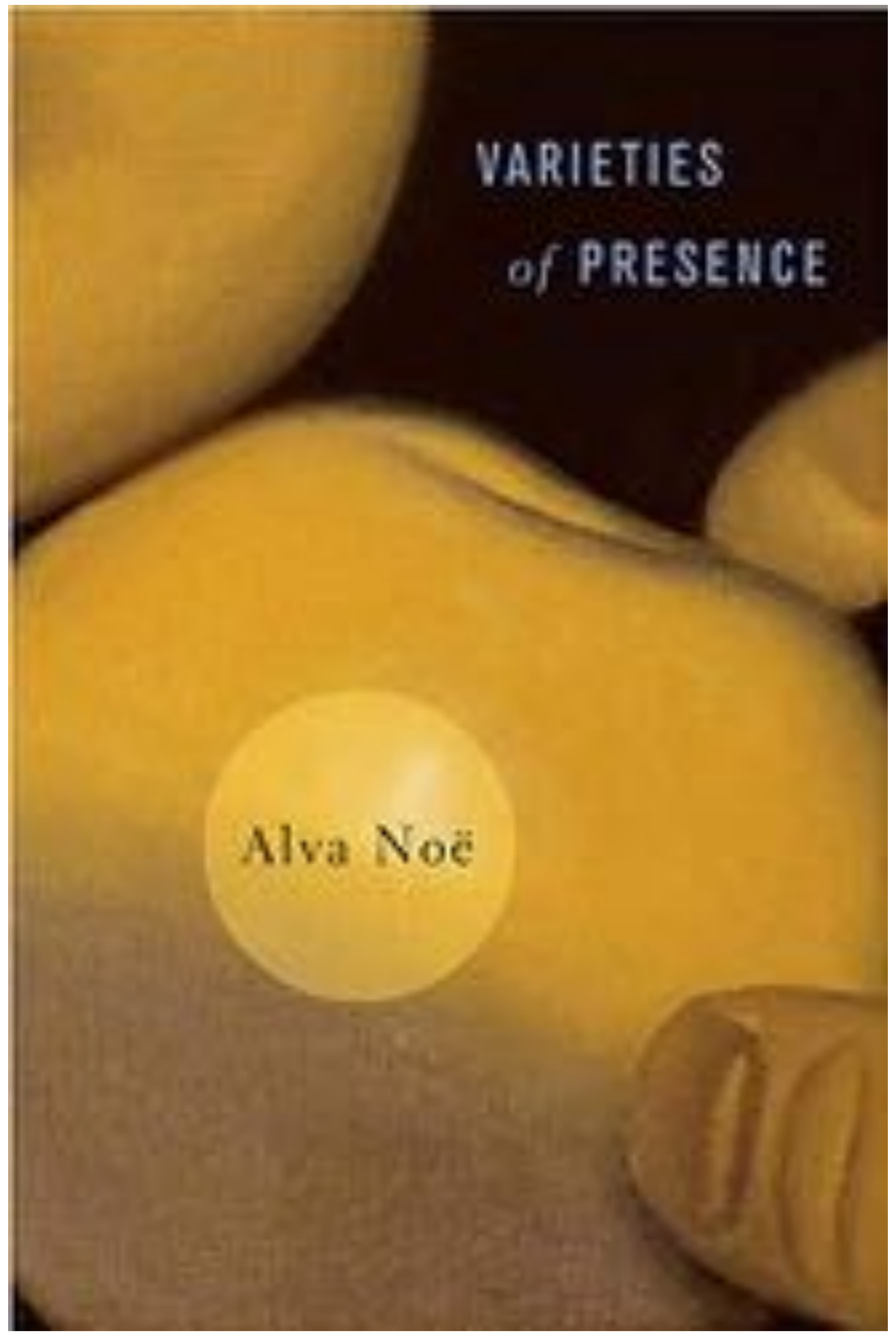
Koray Tahiroğlu
Department of Art and Media, Aalto University, School of ARTS



"Perception is not something that happens to us, or in us,"
"It is something we do."

Perception is not a process in the brain, but a kind of skillful activity of the body as a whole.

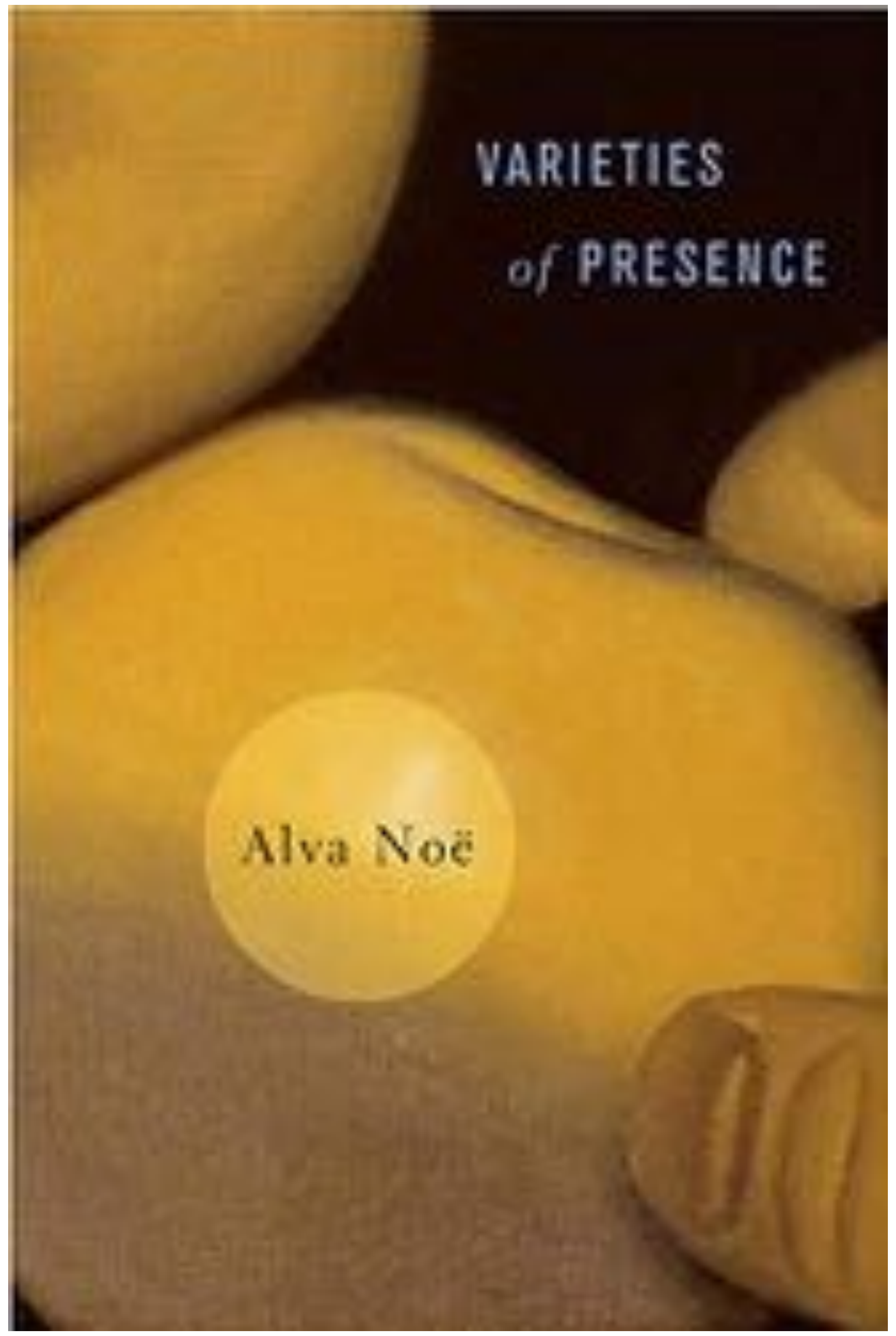
Alva Nöe - Action in Perception, MIT Press



"I begin with an everyday experience. You go to an art gallery and you look at a work of art in an unfamiliar style by an artist you don't know. It happens, sometimes, in a situation like this, that the art work strikes you…."

Perception is not a process in the brain, but a kind of skillful activity of the body as a whole.

Alva Nöe - Varieties of Presence, MIT Press



"I begin with an everyday experience. You go to an art gallery and you look at a work of art in an unfamiliar style by an artist you don't know. It happens, sometimes, in a situation like this, that the art work strikes you…."

"Your experience of art work is transformed."

"Whatever the change has occurred, it enables you now to perceive, in the work, what you couldn't perceive before."

Alva Nöe - Varieties of Presence, MIT Press

Sonic interaction design

Sonic interaction design is the study and exploitation of sound as one of the principal channels conveying information, meaning, and aesthetic/emotional qualities in interactive contexts

Davide Rocchesso, Stefania Serafin, Frauke Behrendt, Nicola Bernardini, Roberto Bresin, Gerhard Eckel, Karmen Franinović, Thomas Hermann, Sandra Pauletto, Patrick Susini, and Yon Visell, (2008). Sonic interaction design: sound, information and experience. In: CHI '08 Extended Abstracts on Human Factors in Computing Systems (Florence, Italy, April 05 – 10, 2008). CHI '08. ACM, New York, NY, 3969–3972.

Sonic interaction design

“there is a tight coupling between auditory perception and action.”

Salvatore M. Aglioti and Mariella Pazzaglia (2010). "Representing actions through their sound". "Experimental Brain Research" 206(2): 141–151. doi:10.1007/s00221-010-2344-x

AHNE - audio haptic navigation environment

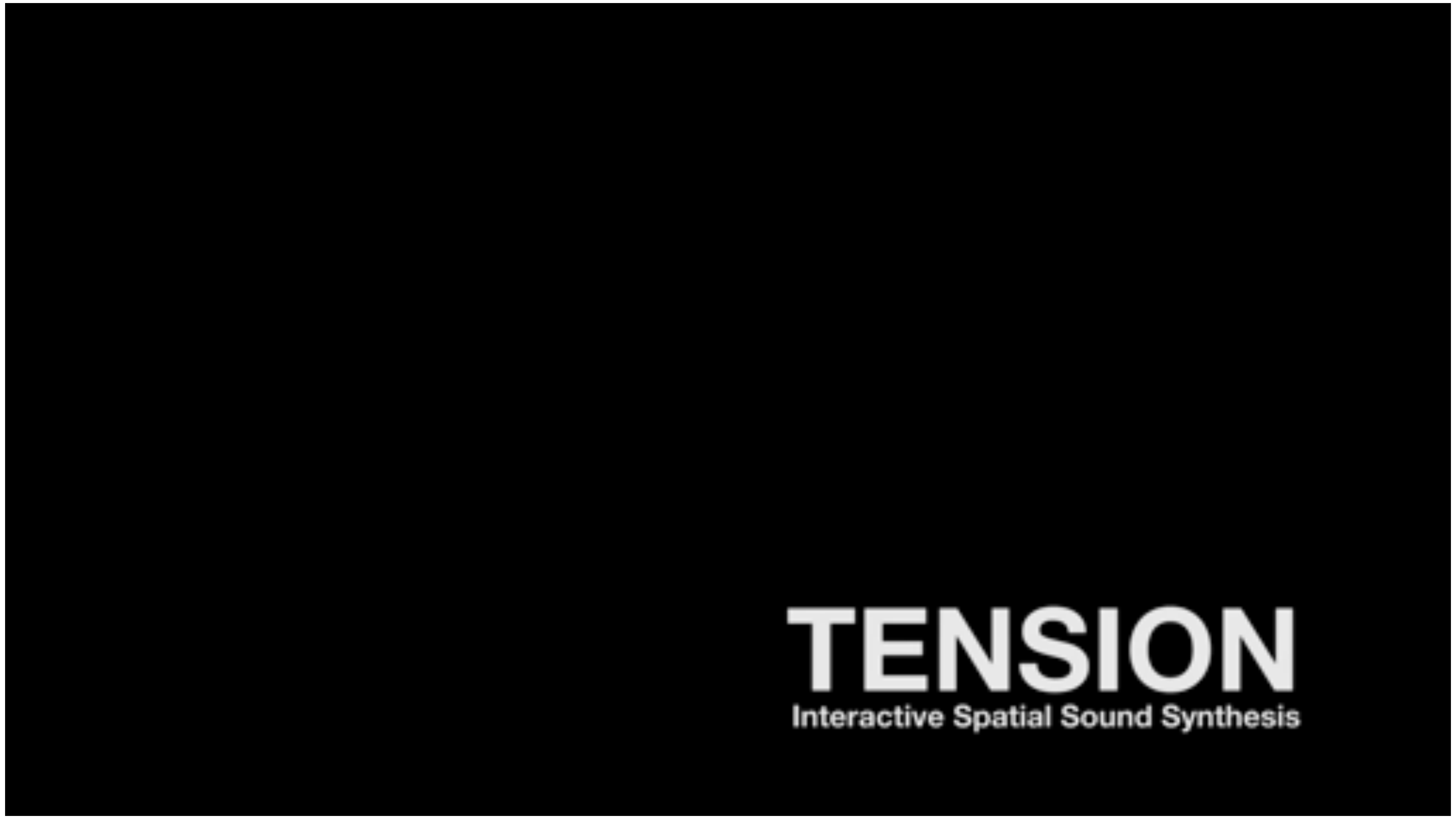
AHNE - audio haptic navigation environment

https://vimeo.com/39778872

Niinimäki, M., & Tahiroglu, K. (2012). AHNE: a novel interface for spatial interaction. In CHI'12 Extended Abstracts on Human Factors in Computing Systems (pp. 1031-1034).

Tahiroğlu, K., Kildal, J., Ahmaniemi, T., Overstall, S., & Wikström, V. (2012, August). Embodied interactions with audio-tactile virtual objects in AHNE. In International Conference on Haptic and Audio Interaction Design (pp. 101-110). Springer, Berlin, Heidelberg.

TENSION



https://vimeo.com/28000196

Niinimäki, M., & Tahiroglu, K. (2012). AHNE: a novel interface for spatial interaction. In CHI'12 Extended Abstracts on Human Factors in Computing Systems (pp. 1031-1034).

Tahiroğlu, K., Kildal, J., Ahmaniemi, T., Overstall, S., & Wikström, V. (2012, August). Embodied interactions with audio-tactile virtual objects in AHNE. In International Conference on Haptic and Audio Interaction Design (pp. 101-110). Springer, Berlin, Heidelberg.

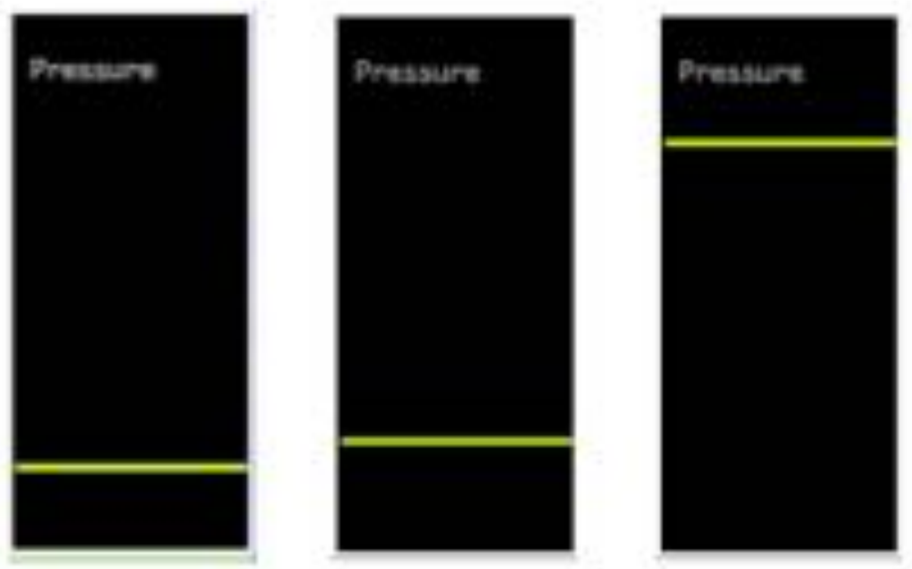
creating an illusion of physicality

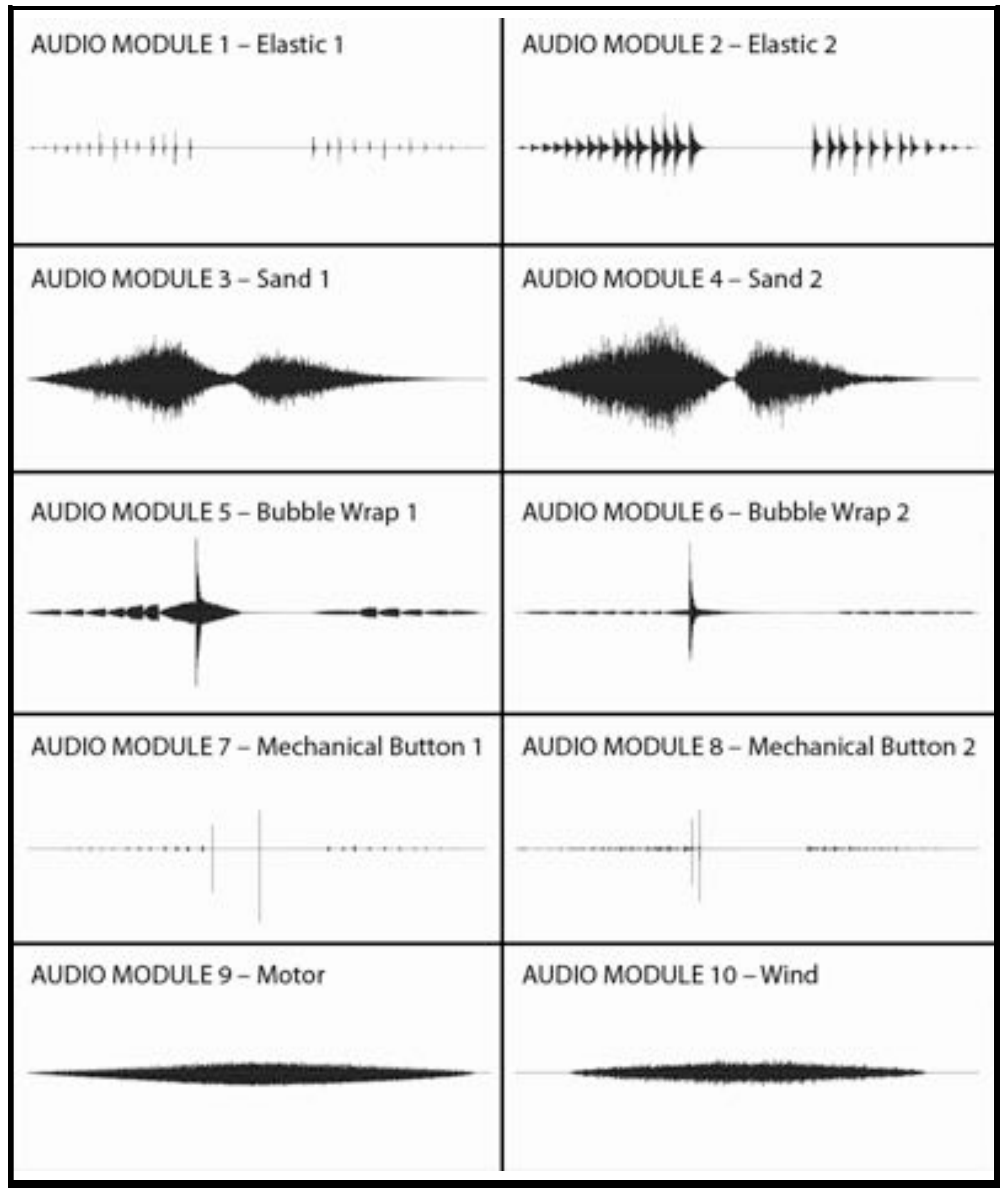
= haptics (tactile) + audio + visual

Lai, C.-H., Niinimaki, M., Tahiroğlu, K., Kildal, J. and Ahmaniemi, T. Perceived Physicality in Audio-Enhanced Force Input. Proc. ICMI'11 ACM, (2011), 287-294.

creating an illusion of physicality

= haptics (tactile) + audio + visual

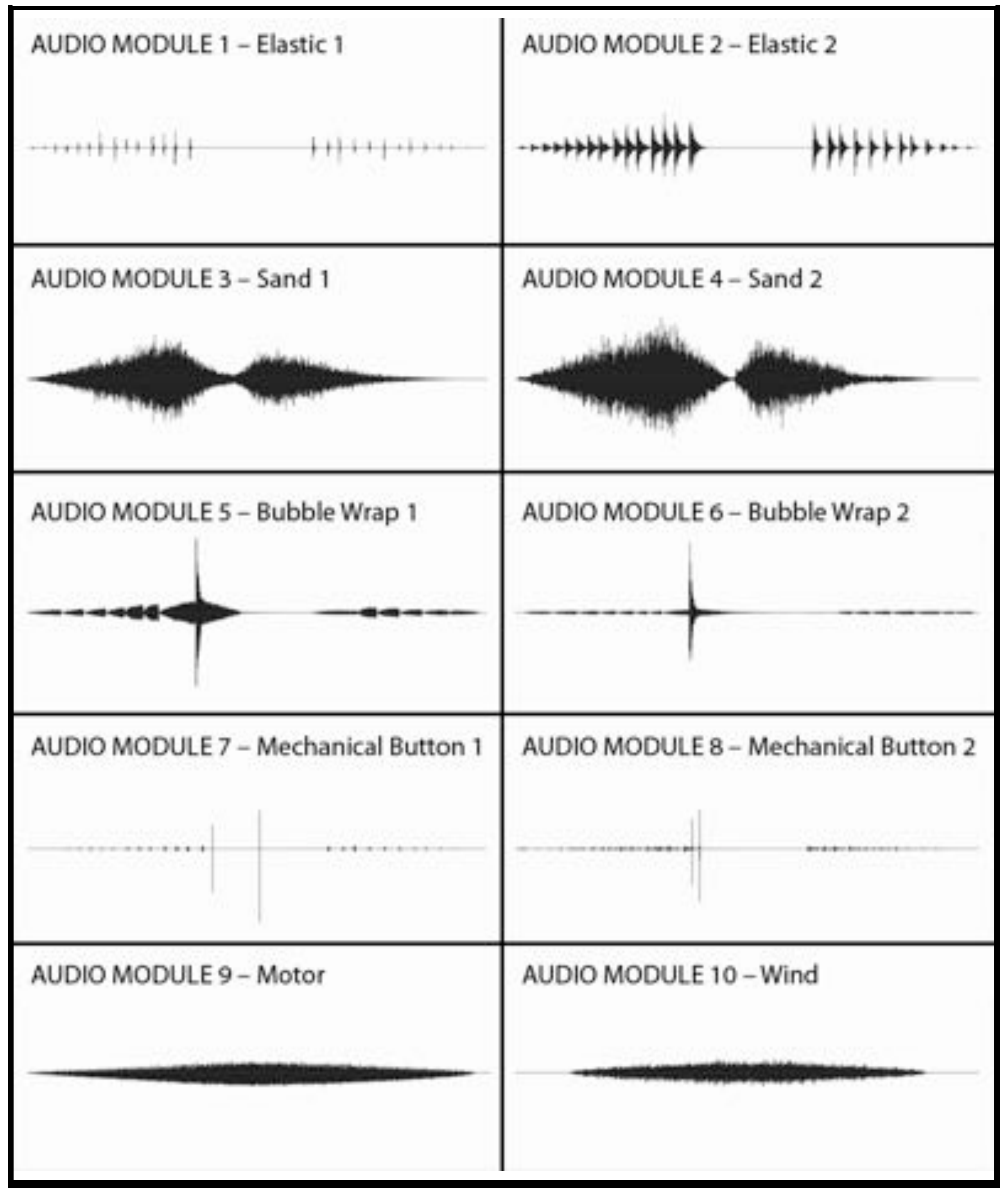




Lai, C.-H., Niinimaki, M., Tahiroğlu, K., Kildal, J. and Ahmaniemi, T. Perceived Physicality in Audio-Enhanced Force Input. Proc. ICMI'11 ACM, (2011), 287-294.

creating an illusion of physicality

= haptics (tactile) + audio + visual



“It fools me that there is a haptic feedback (that's what I think, but it's not really). It's a bit strange, because I hear haptic feedback but there is none.”

“It feels like my finger is moving much more than in any other situation.”

“Some mechanic machine I am controlling. I am kind of accelerating, not directly controlling the speed. There's this lag here. Giving it more force, then it starts rotating faster.”

Lai, C.-H., Niinimaki, M., Tahiroğlu, K., Kildal, J. and Ahmaniemi, T. Perceived Physicality in Audio-Enhanced Force Input. Proc. ICMI'11 ACM, (2011), 287-294.

MARSUI
Malleable Audio Reactive
Shape Retaining
User Interface

https://vimeo.com/61946800

Wikström, V., Overstall, S., Tahiroğlu, K., Kildal, J., & Ahmaniemi, T. (2013). Marsui: malleable audio-reactive shape-retaining user interface. In CHI'13 Extended Abstracts on Human Factors in Computing Systems (pp. 3151-3154).

Sound perception

Definition

“The ability of our hearing system to translate the complex acoustic world into a comprehensible perceptual image” [1]

> Why is the acoustic world complex?

> What are the elements are involved in this process?

[1] Plomp, R. (2001). *The intelligent ear: On the nature of sound perception*. Psychology Press.

Why is the acoustic world complex?

A top-down
approach

A complex acoustic world: The auditory scene

[Listen](#)

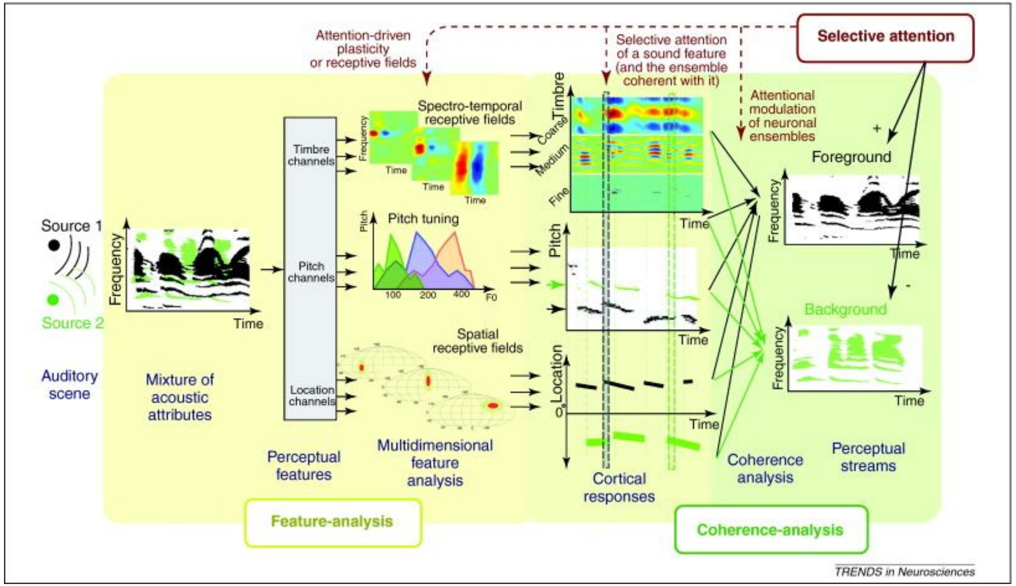
Auditory scenes are compound of many simultaneous auditory objects.

We call auditory object to the “**perception of a physical sound source in the environment and its behavior over time** rather than the source itself or the sounds it emits” [1].

Why is this distinction important?

“Whether a sound is figure or ground has partly to do with acculturation (trained habits), partly with the individual’s state of mind (mood, interest) and partly with the individuals relation to the field (native, outsider). It has nothing to do with the physical dimensions of the sound, for I have shown how even very loud sounds, such as those of the Industrial Revolution, remained quite inconspicuous until their social importance began to be questioned” [2]

Sound properties such as **frequency, intensity, onsets, location** help us to group auditory objects from the same sound source and discern them from other active sound sources.



Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*(3), 114–123. https://doi.org/10.1016/j.tins.2010.11.002

[1] Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research*, *229*(1–2), 225–236. https://doi.org/10.1016/j.heares.2007.01.011

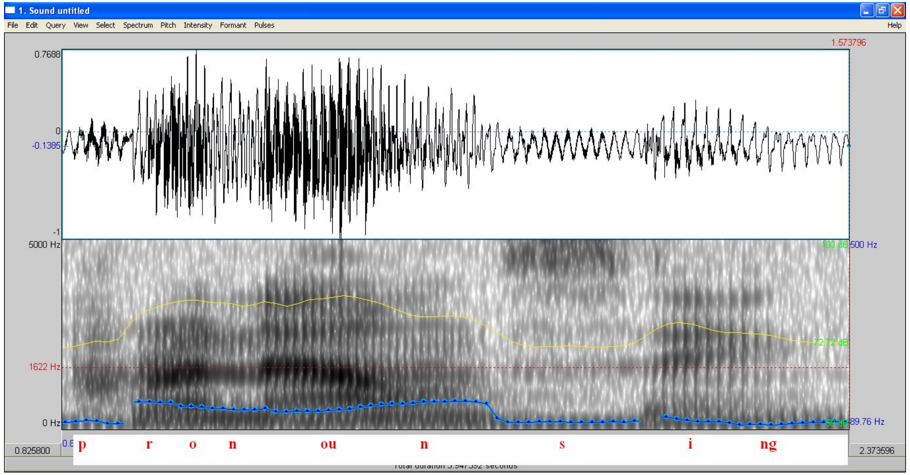
[2] Schafer, R. M. (1994). Perception. In *The Soundscape Our Sonic Environment and the Tuning of the World* (pp. 151–160). Destiny Books.

A complex acoustic world: Sound components

The characteristics of the waveform (evolution of sound over time) determine its timbre.

Sounds can be represented over time or frequency domain

When represented over the frequency domain, we see that sounds can be separated into smaller components called partials, each one with their own frequency, intensity and lifespan.



Waveform and spectrogram of "pronouncing"

A complex acoustic world

Our auditory system not only distinguishes the different auditory objects in a auditory scene.

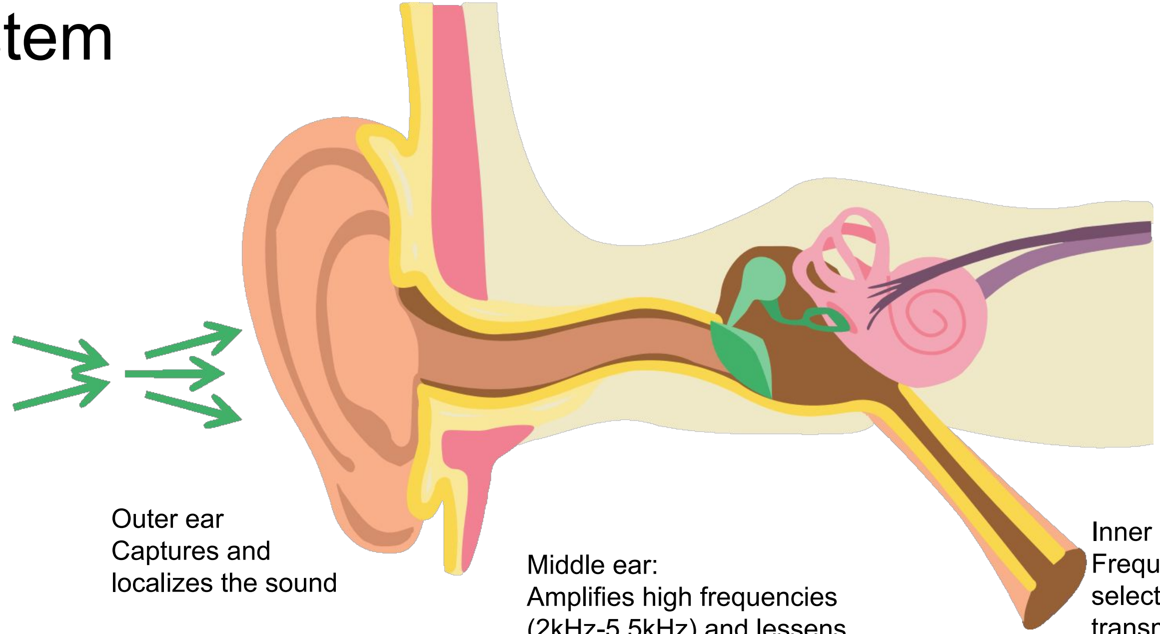
It also separates the different components of sounds to discern different sound sources.

Our auditory system is supported by other sensory input [1]:

- Events in nature stimulate more than one sensory system
- Visual input supports auditory comprehension.
- Our sensory systems work in tandem to adjust our actions depending on the environment.

[1] Warren, R. M. (1982). The Relation of Hearing to Other Senses. In *Auditory perception: A new synthesis* (pp. 188–195). Pergamon Press.

The process: Encoding and decoding in the auditory system



Auditory cortex:
Processing and decoding of sound information

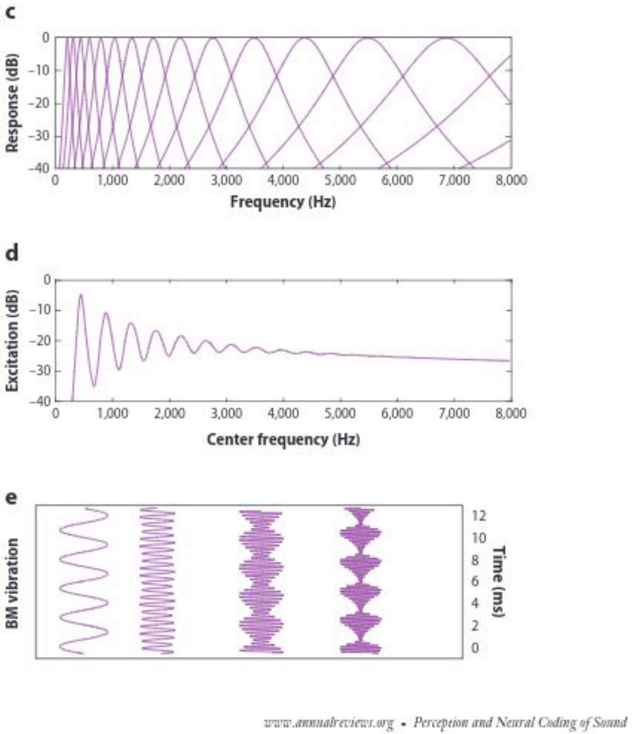
Outer ear
Captures and localizes the sound

Middle ear:
Amplifies high frequencies (2kHz-5.5kHz) and lessens the intensity of strong sound stimuli

Inner ear:
Frequency selectivity and transmission to the auditory cortex

Illustration of how pressure waves reach the ear and propagate throughout auditory system

The process: Band filtering in the inner ear



(*c*) Auditory filter bank representing the filtering that occurs in the cochlea. (*d*) Excitation pattern, or the time-averaged output of the auditory filter bank. (*e*) Sample time waveforms at the output of the filter bank, simulating basilar membrane (BM) vibration, including filters centered at the F0 (440 Hz) and the fourth harmonic (1,760 Hz), illustrating resolved harmonics, and filters centered at the eighth (3,520 Hz) and twelfth (5,280 Hz) harmonics of the complex, illustrating harmonics that are less well resolved and show amplitude modulations at a rate corresponding to the F0. Figure modified with permission from Oxenham (2012).

Oxenham, A. J. (2018). How We Hear: The Perception and Neural Coding of Sound. *Annual Review of Psychology*, *69*, p. 35. https://doi.org/10.1146/annurev-psych-122216-011635

Discrimination: Pitch, loudness & localization

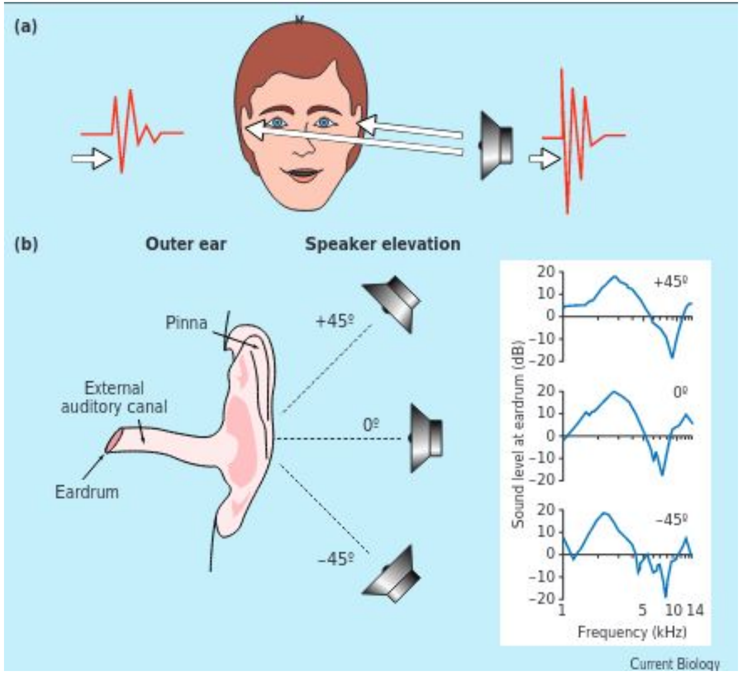
Pitch: the pitch of tones occurring in music and speech is primarily determined by the lower harmonics resolved by the ear. The periodicity of the unresolved higher harmonics may also contribute, but to a lesser extent (Houtsma & Smurzynski, 1990). [1]

Loudness: Auditory intensity. Primarily determined by its sound-pressure level but also by the spectral structure. The more the intensity is spread over a wider frequency range, the louder the tone seems to be [1]. TEST

Simultaneously, A constant level of sound is said to "fatigue" the auditory system and produce a decreased level of sensation, e.g., the loudness of a constant tone falls off with duration [2]

This is also true for other sensory systems. Our sensory input is constantly recalibrated and adapts to what is exposed to [3].

Localization: Conditioned by interaural time, sound level, and filtering of higher frequencies by the body, head and outer ear as well as by familiarity with the sound source or expectations of where the sound may come from. However, our visual input can dominate on our perception of localization [3].



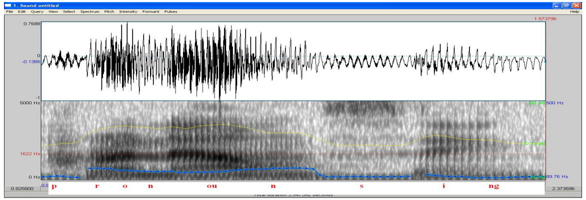
[1] Plomp, R. (2001). *The intelligent ear: On the nature of sound perception*. Psychology Press. pp. 25-29

[2] Truax, B. (2001). *Acoustic communication*. Greenwood Publishing Group.

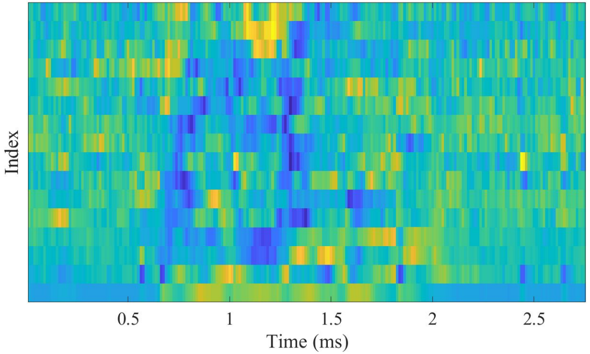
[3] Warren, R. M. (1982). The Relation of Hearing to Other Senses. In *Auditory perception: A new synthesis* (pp. 188–195). Pergamon Press.

Machine listening: FFT feature extractions

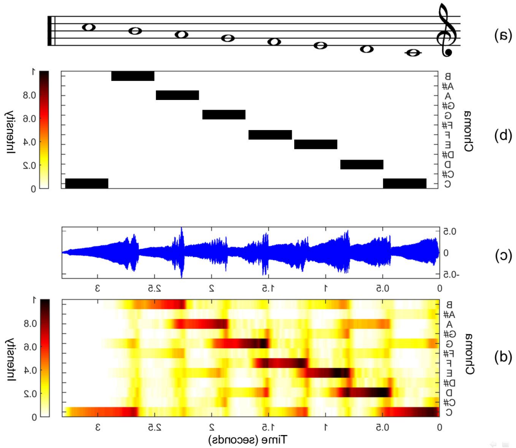
Applying Fourier’s theorem to sound: Any waveform can be described in terms of the amplitude distribution of its harmonics



Waveform and spectrogram of "pronouncing"



MFCC
Source: https://wiki.aalto.fi/display/ITSP/Cepstrum+and+MFCC



Postphenomenology

For phenomenologists, our understanding of sound perception goes beyond our experience with auditory scenes. It is also our interaction with and participation in the acoustic world what conditions our perception of it.

Don Ihde goes a step further and adds that technology mediates the world we live in to what is perceived as science. But because our bodies adapt to different technologies and technological contexts, technology also transforms our perception of the world.

[1] Dourish, P. (2001). *Where the action is*. MIT press Cambridge.

[2] Ihde, D. (2002). *Bodies in technology* (Vol. 5). U of Minnesota Press.

Further reading:

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*(10), 465–471. https://doi.org/10.1016/j.tics.2004.08.008

Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology*, *20*(3), 361–366. https://doi.org/10.1016/j.conb.2010.03.009

RADER, C. M., & MALING, G. C. (1967). What Is the Fast Fourier Transform? *Proceedings of the IEEE*, *55*(10), 1664–1674. https://doi.org/10.1109/PROC.1967.5957

Pedersen, P. (1965). The Mel Scale. In *Source: Journal of Music Theory* (Vol. 9, Issue 2). Winter. https://about.jstor.org/terms

SHORT INTRO TO PHONETICS

Phonetics & Phonology

- **Phonetics**: the study of speech production, speech signals, and speech perception
  - *Acoustic, articulatory, and auditory aspects of speech sounds*
  - *Measurable properties*

- **Phonology**: the study of sound systems in language
  - *Abstraction of speech units to study possible combinations of sound*
  - *Classification*

Written language models spoken language elements, but spoken language is continuous

Elementary units of spoken language

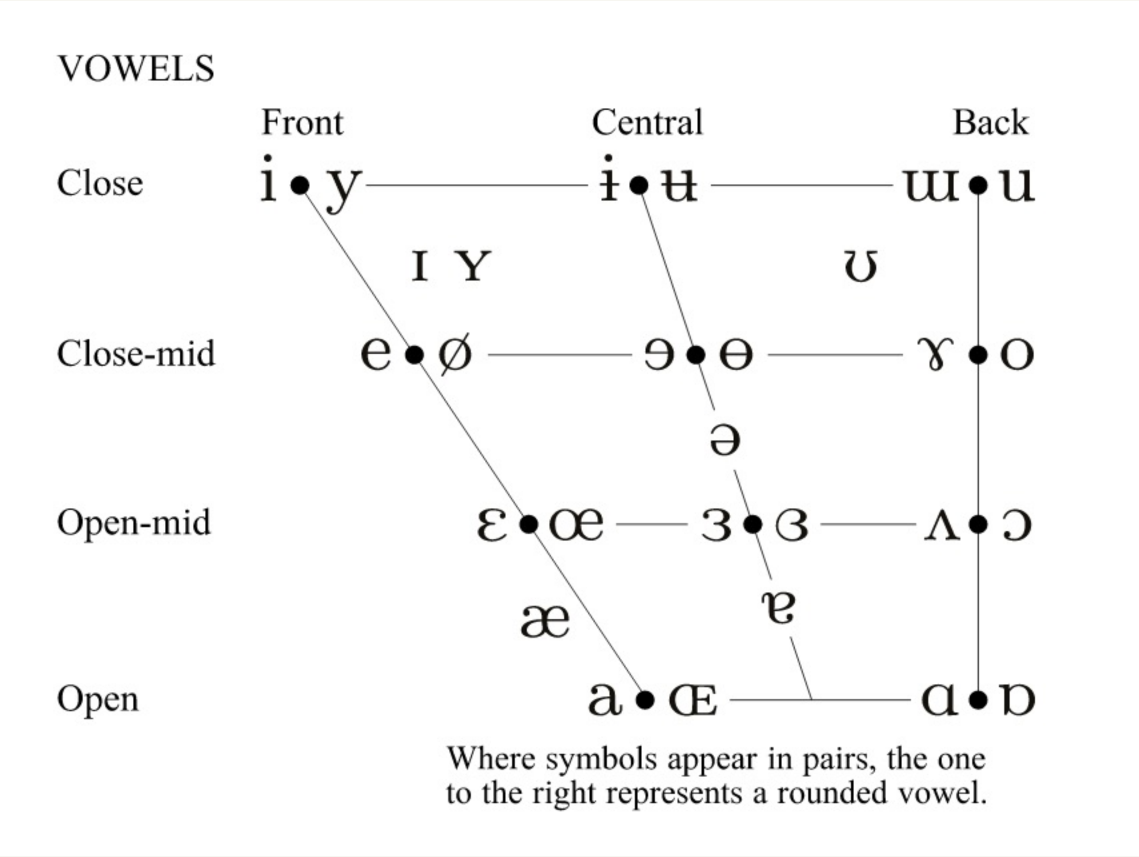
Speech consists of elementary units of increasing time-scale

- ***Phones:*** *individual speech sounds, notation e.g. [k], [e]*
  - Segmental units, often non-meaningful alone, but create and distinguish meaning in sequences
  - Each phone is associated with an **articulatory gesture** (production) and **acoustic cue** (perception)
  - Divided into vowels and consonants (and semi-vowels…)
- ***Syllables:*** *sequences of phones*
- ***Words:*** *sequences of syllables*

- **Phoneme:** an abstraction for a speech sound which distinguishes meaning, notation e.g. /k/, /e/
  - Two words which differ in one phoneme and have different meanings are a **minimal pair**, e.g. "*cat*" [kæt] and "*bat*" [bæt]
  - Phones which can be used interchangeably without changing the meaning of a word are **allophones** (realizations of the same phoneme)

Vowel classification

- Openness/closedness
- Frontness/backness
- Lip rounding



Consonant classification

- Place of articulation
- Manner of articulation

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

Syllables → words → utterances

- **Syllables:** sequences of phones
  - *Onset, nucleus, coda, e.g. ”kissa” → kis-sa; in the first syllable, [k] is the onset, [i] is the nucleus, [s] is the coda*
  - *Rhythmic units*
  - *Sonority increases towards the nucleus of the syllable*
  - *Sonority hierarchy in speech sounds is typically vowel > semi-vowel > liquid > nasal > obstruent*
- **Words:** smallest meaning-bearing units
- **Utterances:** an unpaused act of speech, made up of one or more words or fillers

Prosodic properties of speech

- **Intonation:** variation in the fundamental frequency, perceived in terms of pitch
  - *Used to signal e.g. emphasis, a question, surprise*
- **Stress:** relative emphasis on a syllable or set of syllables within a phrase (or within a word, when talking about **word stress**)
- **Rhythm:** a combination of stress, pause, duration, syllable structure; recurrence of patterns in speech

https://wiki.aalto.fi/display/ITSP/Linguistic+structure+of+speech