

## Instrumental and vocal timbre perception

Prof. Caroline Traube

Laboratoire d'informatique, acoustique et musique (LIAM)  
Faculté de musique, Université de Montréal, Québec, Canada

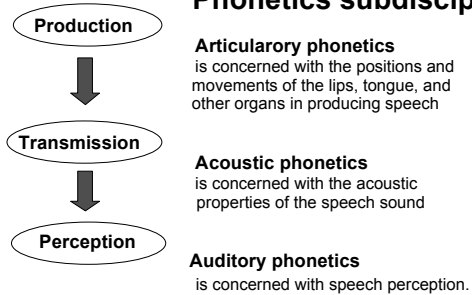
## Vocal timbre perception

### The study of speech sounds Phonetics and phonology

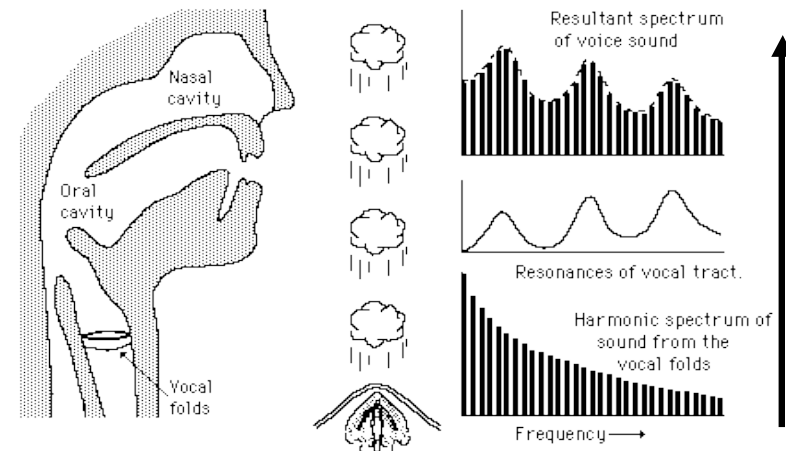
**Phonetics** is the scientific study of the sounds of language and of the spoken communication process. Phoneticists are more concerned with the sounds of speech than the symbols used to represent them.

**Phonology** is the study of the function of phonemes in a given language and the opposition and contrasting relations in the system formed by the sounds of this language.

#### Phonetics subdisciplines



### Voice acoustics



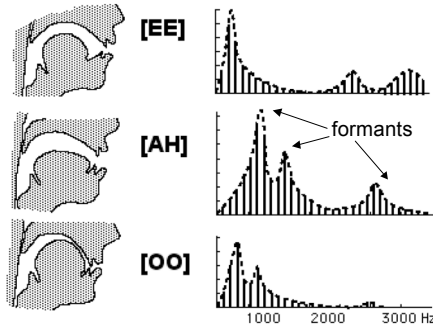
## Timbre of vowels : the role of resonators

The timbre of a vowel depends on :

-The **number of active resonators** (among oral, labial and nasal cavities)

- The **shape of the oral cavity** determined by the position of the tongue in the mouth (front/back, high/low)

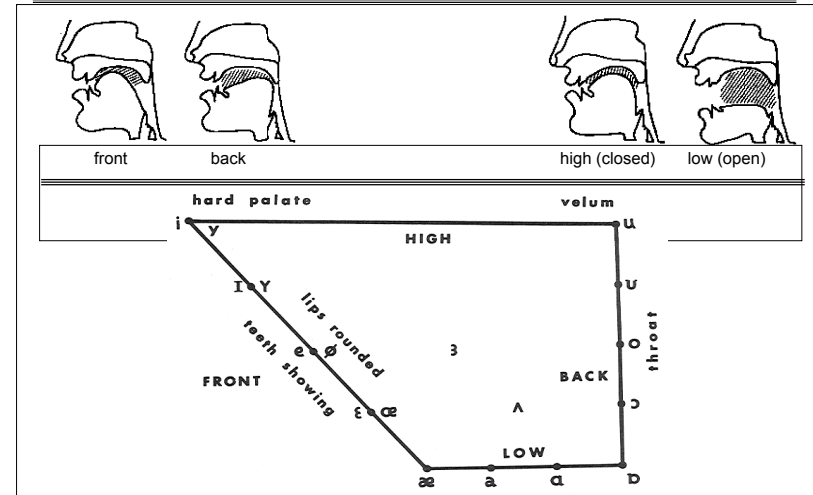
- the **volume of the oral cavity** depending on the degree of opening of the jaws.



The central frequency of formants depends on the configuration of the articulators (tongue, lips, palate).

The central frequency of a formant does not necessarily correspond to the frequency of a harmonic component !

## Timbre of vowels : the role of resonators



## Spectral envelope of a vowel

The **spectral envelope** connects the summits of the magnitude spectrum components.

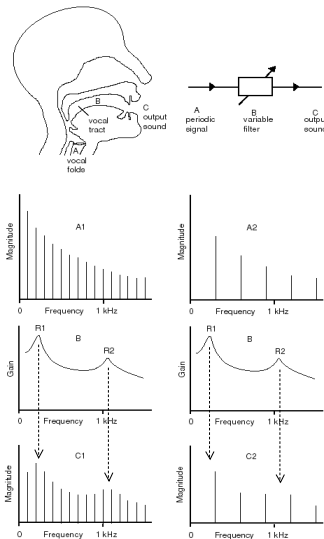
In the case of vowels :

**resonances** in the vocal tract  
→ **formants** (region of the spectrum where the energy is concentrated)

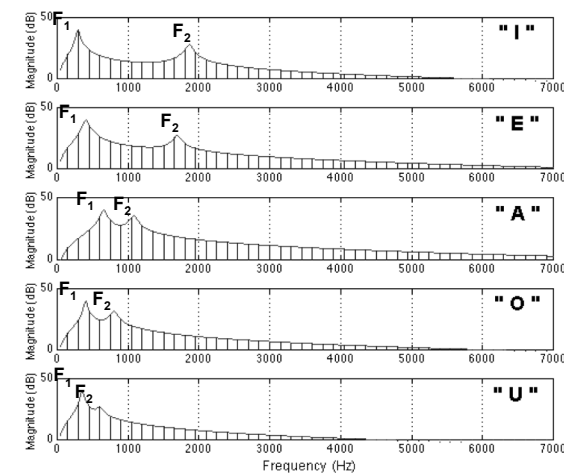
Typically, a vowel has 5 main formants.

For a given vowel, **the position of the spectral envelope along the frequency axis is absolute, regardless of the fundamental frequency.**

Therefore, a timbre-preserving transposition is not just a stretching or compression of the spectral envelope.

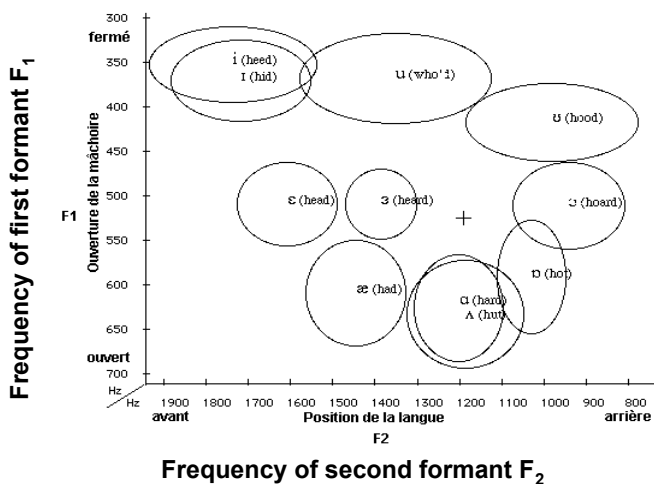


## We can recognize a vowel from just two formants

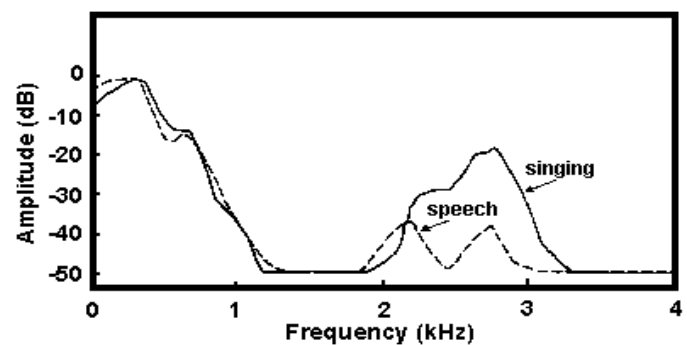


**F<sub>1</sub>** = first formant  
**F<sub>2</sub>** = second formant

## Vowel space ( $F_1$ - $F_2$ plane)



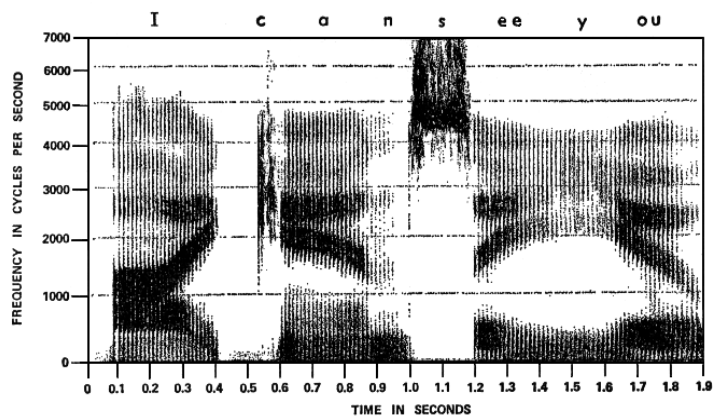
## The singer's formant



The singer's formant appears around 3000 Hz and is in fact the reunion of several formants (3d, 4th and 5th).

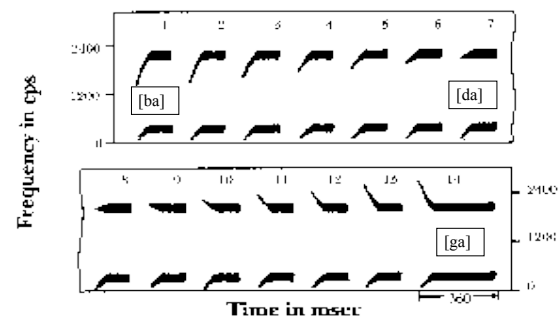
It allows the singer to be better heard when singing over an orchestra (otherwise the orchestra would easily mask the singer).

## Spectrogram of speech



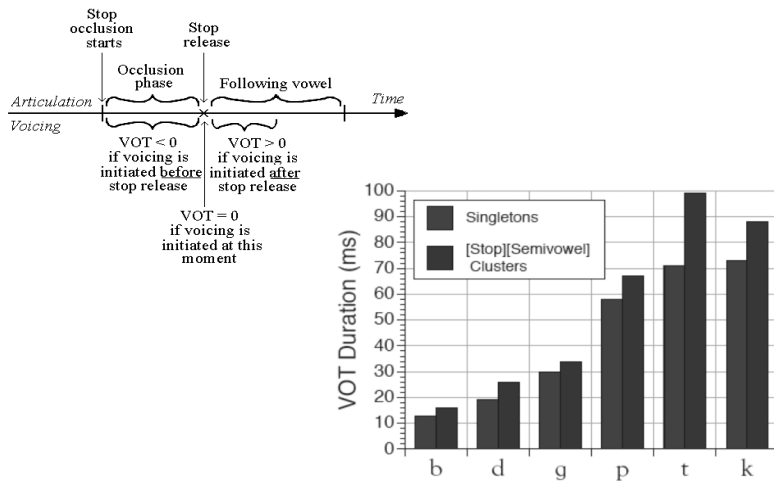
On a spectrogram, formants appear as darker zones. We can clearly see « formants transitions » corresponding to diphthongs (very frequent in English).

## Plosive consonants and formant transitions



This demonstration illustrates the **categorical perception** of consonants.

## Plosive consonants and Voice Onset Time (VOT)



## Distinctive features of speech sounds

The distinctive feature theory was proposed by Jakobson, Fant and Halle in 1951 and then later revised and refined by Chomsky and Halle in 1968.

The theory codifies certain long-standing observations of phoneticians by hypothesizing that many sounds of speech can be categorized based on the presence or absence of certain distinctive features: whether the mouth is open, whether there is a narrowing of the vocal tract at a particular place, whether a consonant is aspirated.

Those properties are the features that characterize and distinguish the phonetic content of a language.

The theory can be applied, with only slight modifications, to all human languages throughout the world. Jakobson, Fant and Halle detected twelve inherent distinctive features in the languages of the world.

## Physiological description of speech sounds

The principal physiological factors that are considered when distinguishing vowels from one another are :

- Movement of the **tongue forward or backward** with the jaw held steady.  
Example: (English) *panned - pond - pawned*.
- Movement of the mouth and jaw **from almost closed to fully open** with the tongue held steady.  
Example: (English) *mean - mane - man*
- **Rounding or non-rounding** of the lips with the tongue and jaw held steady.  
Example: (German) *Tür-Tier*
- **Opening or closing** the passage to the **nasal cavity** with the tongue and jaw held steady.  
Example: (French) *bon-beau*

**Consonants** differ according to the following principal criteria:

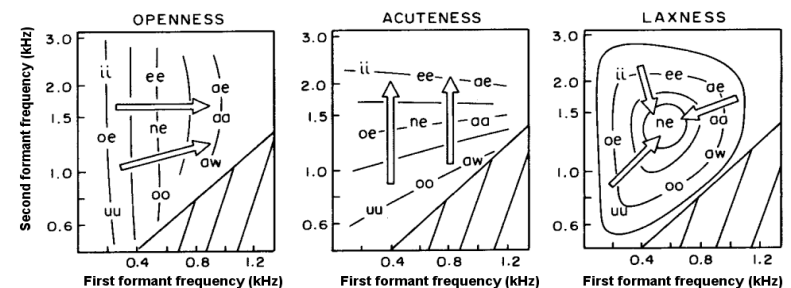
- Presence or absence of **voicing** (vocal folds vibration). Example: *din - tin*
- Complete or partial **obstruction of the air flow**. Example: *tin - sin*
- Closure or non-closure of the **velum**. Example: *dip - nip*
- Surmounting or circumventing the obstruction. Example: *rip - lip*

## Slawson's equal-value contours

In his book *Sound Color* (1985), Slawson addresses the following question: "How can one aspect or dimension of sound color be held constant as other dimensions of sound color are varied?"

He answers by first designating three of the distinctive features of vowels (compactness [d.f. 7], acuteness [d.f. 9] and laxness [d.f. 12]) as candidates from which to derive dimensions of sound colour.

Then he determines **equal-value contours** for these distinctive features.



## Distinctive features of sound color

- **OPENNESS** (replacing the term COMPACTNESS) is named for the tube shape with which it is correlated. The approximate acoustic correlate of OPENNESS is the frequency of the first resonance.

- **ACUTENESS** reflects its connotation of high or bright sound. It increases with increasing frequency of the second resonance.

- **LAXNESS** is said to correspond to a relatively relaxed state of the articulatory musculature. The equal LAXNESS contours are closed curves on the (F1, F2) plane centered on the maximally lax point. This central point correspond to the formant values that would arise, in theory, from the vocal mechanism in the position to which it is automatically brought just before beginning to speak (Chomsky, 1968).

The neutral position of the vocal tract can be best approximated by a single tube closed at one end. Since a tube of length L closed at one end can only resonate at frequencies for which L is an odd multiple of one quarter wavelength and since the average length of the vocal tract of males is about 17.5 cm, the resonances appear at approximately 500, 1500, 2500 Hz, etc.

A tense vowel displays a greater deviation from the neutral formant pattern.

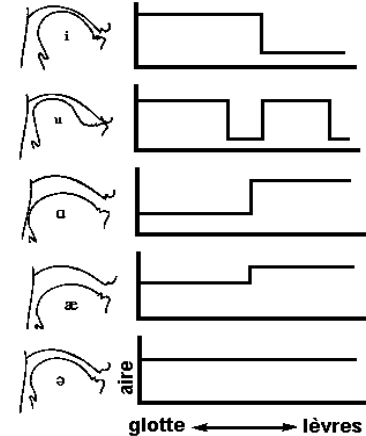
## Instrumental timbre perception

## Motor theory of speech perception

The "motor theory" of speech perception was proposed by A. M. Liberman and his colleagues (Liberman 1967, 1985).

In its most recent form, the model claims that "the objects of speech perception are the intended phonetic gestures of the speaker, represented in the brain as invariant motor commands that call for movements of the articulators through certain linguistically significant configurations". In other words, we perceive the articulatory gestures the speaker intends to make when producing an utterance (Moore 1997).

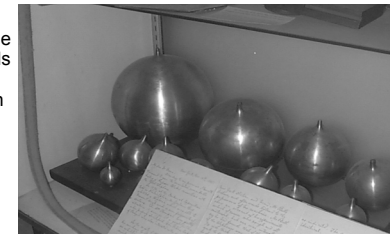
A second claim of the motor theory is that there is an **intimate and innate link between speech perception and speech production**. Perception of the intended gestures occurs in a specialized speech mode whose main function is to automatically convert an acoustic signal into an articulatory gesture.



## Definitions of timbre through history

### Helmholtz (1877)

Helmholtz disregards all irregular portions of the motion of the air, and the mode in which sounds commence or terminate, directing his attention solely to the « musical part » of the tone, which corresponds to a uniformly sustained and regularly period motion of the air. (*On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 1877)



« Quality of tone (*Klangfarbe*) should depend upon the manner in which the motion is performed within the period of each single vibration. »

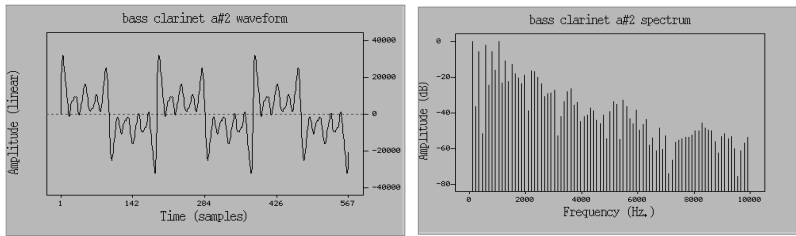
« ...differences in musical quality of tone depend solely on the presence and strength of partial tones, and in no respect on the differences in phase under which these partial tones enter into composition. It must be here observed that we are speaking only of musical quality as previously defined. »

### Stumpf (1890)

Stumpf listed no less than 20 relevant semantic scales as wide-narrow, smooth-rough, round-sharp, etc., concluding that « this wealth of adjectives is comparable only with those used by wine merchants for extolling their products ».

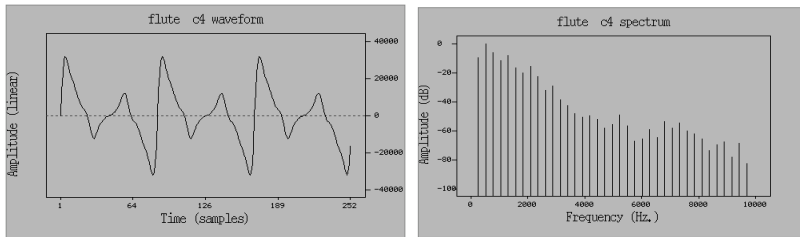
Signal : clarinet A#2

→ magnitude spectrum

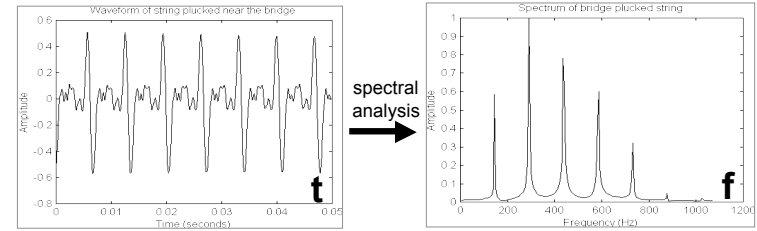


Signal : flute C4

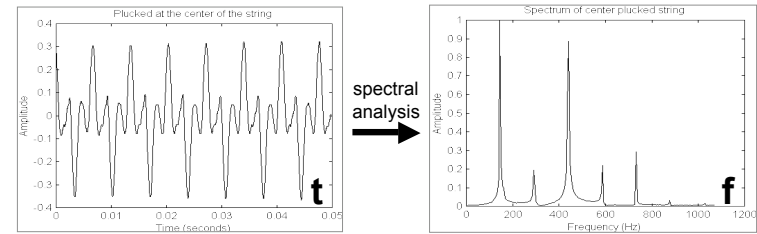
→ magnitude spectrum



Guitar: string plucked by the bridge

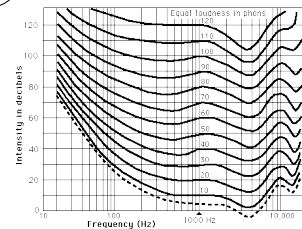
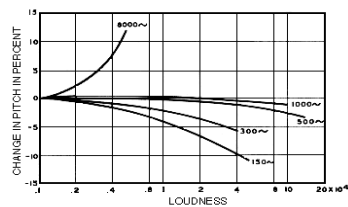
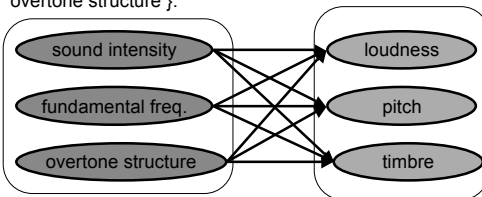


Guitar: string plucked by the middle of the string



## H. Fletcher (1934)

Experiments show that a simple one-to-one relationship does not exist between the two sets { loudness, pitch, timbre } and { sound intensity, fundamental frequency and overtone structure }.



« Timbre is that characteristic of sensation which enables the listeners to recognize the kind of musical instrument producing the tone, that is, whether it is a cornet, a flute or a violin. »

« Large changes in loudness or pitch, without in any way changing the overtone structures, will also produce changes in timbre. »

*Loudness, pitch and the timbre of musical tones and their relation to the intensity, the frequency and the overtone structure. Journal of the Acoustical Society of America 6, pp. 59-69 (1934).*

## Seashore (1938)

« Tone quality has two fundamental aspects, namely,

- (1) timbre, which is the simultaneous presence or fusion of the fundamental and its overtones at a given moment, and
- (2) sonance, the successive presence or fusion of changing timbre, pitch, and intensity in a tone as a whole.

The first may be called simultaneous fusion; the second, successive.

## ANSI (1960)

Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar.

Note : timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus.

### Effect of spectrum on timbre

Instruments tones reconstructed one partial at a time

Auditory Demonstrations : no 28 (track 53)

Sound 1	Sound 2
1) 251 Hz (drone) 2) + 501 Hz (fundamental) 3) + 603 Hz, 750 Hz (min. third, fifth) 4) + 1005 Hz (octave) 5) + 1506 Hz 6) + 2083 Hz 7) + 2421, 2721 Hz (next two partials) 8) + all remaining partials	1) 251 Hz (fundamental) 2) + 502 Hz (H2) 3) + 753 Hz (H3) 4) + 1004 Hz (H4) 5) + 1255 Hz, 1506 Hz (H5, H6) 6) + 1757 Hz, 2008 Hz (H7, H8) 7) + 2259 Hz, 2510 Hz, 2761 Hz (H9, H10, H11) 8) + all remaining partials
bell → pseudo-harmonic spectrum	guitar → harmonic spectrum

## Schaeffer (1966)

The perceived timbre is a synthesis of the variations of the harmonic content and of the dynamic evolution of the sound.

In 1948, Pierre Schaeffer realized the experiments known as the "sillon fermé" (*closed groove*) and the "cloche coupée" (*cut bell*).

- (1) le sillon fermé : in looping a sound on itself, Pierre Schaeffer isolates the sound from "what was before it and what will follow it".
- (2) la cloche coupée : by removing the attack of a bell sound, he obtains a flute sound !

He concludes that timbre is not only determined by the overtone structure.

Schaeffer defines the attack timbre, the dynamic timbre and the harmonic timbre (*Traité des Objets Musicaux*, 1966).

## Schouten (1968)

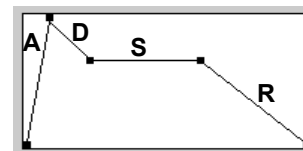
Loudness, pitch and duration are the easiest to ascertain in the overall impression of any sound. For all other qualities we have scarcely more at our disposal than the one and all embracing term: timbre.

A very vague way of brining all other unresolved attributes under one general heading. This is an extremely disappointing state of affairs...

... The vague heading "timbre", though, is precisely the one which covers those invariant acoustic properties which make us recognize a violin for example.

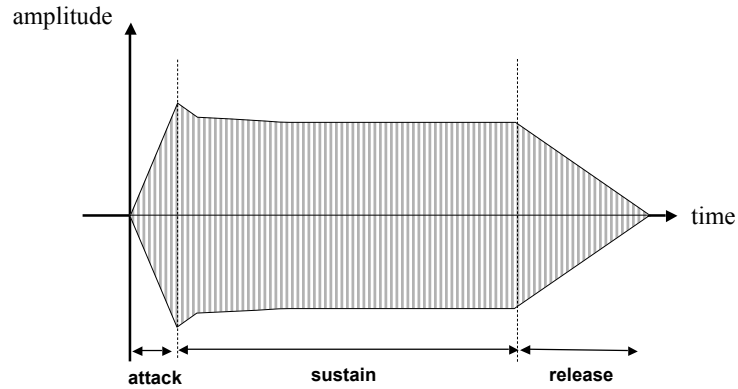
According to Schouten, timbre can be expressed in terms of at least 5 major parameters:

1. The range between tonal and noise-like character,
2. The spectral envelope,
3. The time envelope in terms of rise, duration and decay,
4. The change both of spectral envelope (formant glide) or fundamental frequency micro-intonation),
5. The prefix, an onset of a sound quite dissimilar to the ensuing lasting vibration.



The temporal envelope is the evolution of amplitude through time. For a classical musical object, one can usually distinguish an **attack**, a fast **decay**, a **sustain** and a **release** (ADSR envelope).

## The three phases of an instrumental sound



**Demonstration :**  
effect of tone envelope on timbre

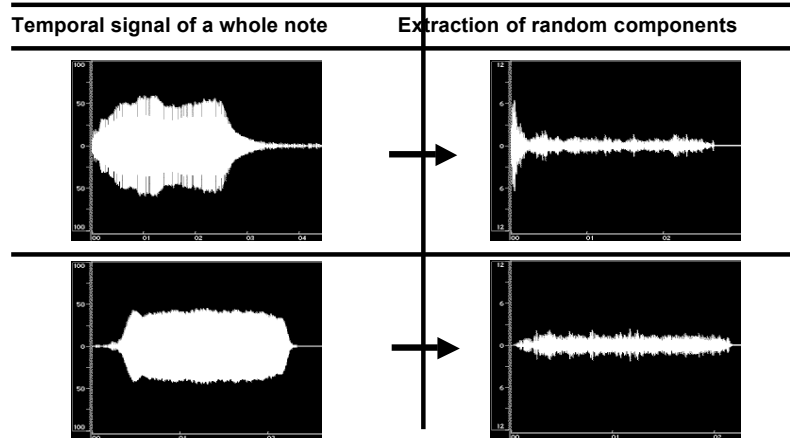
**Auditory Demonstrations (ASA) : no 29 (track 54)**

- 1- recording of a J.C. Bach piano piece
- 2- recording of same piece played backwards
- 3- tape of last recording is played backwards  
(piece is heard forward again but individual notes are reversed)

In the third recording, perceived timbre is not the one of a piano, but rather the one of an organ.

**Mixed spectrum** : combination of harmonic and noisy characters.

It is the case of musical instruments.



## Erickson (1975)

Clearly timbre is a multidimensional stimulus: it cannot be correlated with any single physical dimension.

Analysis of the gamut of clarinet tones might lead one to say that it is three instruments, rather than one.

## Grey (1975)

Timbre may refer to the features of tone which serve to identify that a musical sound originates from some particular instrument or family of instruments, for example, that it is an oboe, or perhaps some sort of double-reed instrument, or maybe just some woodwind instrument.

## Roederer (1975)

Timbre perception is just a first stage of the operation of tone source recognition (in music, the identification of the instrument).

From this point of view, tone quality perception is the mechanism by means of which information is extracted from the auditory signal in such a way as to make it suitable for:

- (1) storage in the memory with an adequate label of identification (learning or conditioning - ex : child learning what a clarinet is), and
- (2) comparison with previously stored and identified information (conditioned response to a learned pattern)

On the other hand, if we listen to a 'new' sound, e.g., a series of tones concocted with an electronic synthesizer, our information-extracting system will feed the cues into the matching mechanism, which will then try desperately to compare the input with previously stored information.

- If this matching process is unsuccessful, a new storage 'file' will eventually be opened up for this new, now identified, sound quality.
- If the process is only partly successful, we react with such judgments as 'almost like a clarinet' or 'like a barking trombone.'

### **Demonstration : change of timbre with transposition Auditory Demonstrations (ASA) : no 30 (track 57)**

Usually, low and high notes played on a musical instrument do not have the same relative spectrum.

For example, a low piano tone contains a small amount of energy at the fundamental frequency. Most of the energy is found in the higher order harmonics.  
On the other hand, a high piano tone typically has a strong fundamental and weaker partials.

In the demonstration, a scale on three octaves is played on a bassoon.  
Then is presented a scale obtained by stretching in time the highest note, in order to produce the frequency of each note (segments from stationary part are removed to maintain the initial duration of notes).

## Plomp (1976)

The ear is not as « phase deaf » as had been suggested by earlier investigators.

... we may conclude that the effect on timbre of varying the phase spectrum of a complex tone is small compared with the effect of varying the amplitude spectrum.

Timbre is determined by the absolute frequency position of the spectral envelope rather than by the position of the spectral envelope relative to the fundamental frequency.

Bismarck found that sharpness as the major attribute of timbre is primarily related to the position of the loudness centre on an absolute frequency scale rather than to a particular shape of the spectral envelope. . . .

The dependence of timbre upon frequency would imply that simple tones are also characterized by a specific timbre, to be distinguished from their pitch. Low frequency tones do indeed sound dull and high-frequency tones sharp... The observation that simple tones have some resemblance, depending upon their frequency, with particular vowels also supports this view. Subjects appear to be able to label simple tones rather well in terms of vowels... This resemblance is related to the frequency of the most characteristic formant or combination of formants.

## Rash and Plomp (1982)

Subjectively, timbre is often coded as the function of the sound source or of the meaning of the sound. We talk about the timbre of certain musical instruments, of vowels, and of sounds that signify certain events in our environment (apparatus, sounds from nature, footsteps, the slapping of a door, etc)

Temporal characteristics of the tones may have a profound influence on timbre :

- Onset effects
  - rise time
  - presence of noise or inharmonic partials during onset
  - unequal rise of partials
  - characteristic shape of rise curve
- Steady state effects
  - vibrato
  - amplitude modulation
  - gradual swelling
  - pitch instability

These characteristics are important factors in the recognition and, therefore, in the timbre of tones.

## Dowling & Harwood (1986)

Timbre (tone color) has always been the miscellaneous category for describing the psychological attributes of sound, gathering into one bundle whatever was left over after pitch, loudness, and duration had been accounted for.

The psychological attributes clustered under the heading timbre fall along more than one psychological dimension; that is, sounds do not simply differ in how much timbre they have. And there are several physical dimensions whose variation causes changes in timbre that interact with each other in complex ways.

## Houtsma (1989)

**Interpretation of the ANSI (1960) definition** : According to this definition, timbre is the subjective correlate of all those sound properties that do not directly influence pitch or loudness. These properties include the

- sound's spectral power distribution
- its temporal envelope
- rate and depth of amplitude or frequency modulation
- degree of inharmonicities of its partials.

## Bregman (1990)

When we do find a characteristic of sound that can be obtained on different instruments, such as vibrato, the characteristic tends to be given a label and no longer falls into the nameless wastebasket of « timbre ».

## Cho, Hall et Pastore (1993)

Timbre is the subjective attribute of source (instrument) that is based on invariant properties that uniquely characterize the tones produced by the source.

Unfortunately, the pursuit of an adequate definition of timbre is both related to and dependent upon establishing which characteristics (or combination of characteristics) are important for perceptually determining an instrument's distinctive sound quality

## Krumhansl (1989)

Different levels of timbral description:

1. « the expressive variations available to performing musicians »
2. « commonalities shared by all oboe tones, all bowed violin tones, all timpani tones, and so on »
3. Broader family distinctions or method-of-production distinctions: « percussive instruments, whose behavior is determined completely at the instant when they are set into motion, and instruments, such as blown and bowed instruments, whose behavior is controlled continuously. »

Alternative set of distinctions for describing sound (Schaeffer, McAdams):

1. « varying degrees of temporal extent or musical complexity. . . . single, discrete sound events that are heard as being produced by a single source. »
2. « emergent properties, such as texture, density, streams, and musical gestures. »
3. « larger-scale musical forms or organizations that grow out of the sound material. »

## Handel (1995)

One possibility is that timbre is perceived in terms of the actions required to generate the event. . . . The perception of the production invariances would allow us to hear the same object in spite of large changes in the acoustical signal.

Another possibility is that timbre is perceived simply in terms of the acoustic properties and that the connection between the acoustic properties and the object is learned by experience. In this view, the acoustic properties are used to figure out what event was most likely to have produced that sound.

## Summary : physical parameters related to timbre

- temporal envelope
- spectral envelope
- absolute frequency position of the spectral envelope
- variations of harmonic contents
- position of spectral centroid -> brightness or sharpness
- harmonic and noise components ratio
- inharmonicity ratio
- odd/even harmonic ratio
- synchronicity of partials
- onset effects
  - rise time
  - presence of noise or inharmonic partials during onset
  - unequal rise of partials
  - characteristic shape of rise curves
- steady state effects
  - vibrato
  - amplitude modulation
  - gradual swelling
  - pitch instability

## An important contribution to understanding timbre perception :

### the Multidimensional Scaling Study of J.M. Grey

Several studies have performed multidimensional scaling analyses on dissimilarity ratings for musical instrument tones or synthesized tones with characteristics that resemble those of musical instruments (Plomp, 1970, 1976; Weddin et Goude, 1972; Wessel, 1973; Miller et Carterette, 1975; Grey, 1977; Krumhansl, 1989.) In all of these studies, the perceptual axes have been related either qualitatively or quantitatively to acoustic properties of the tones.

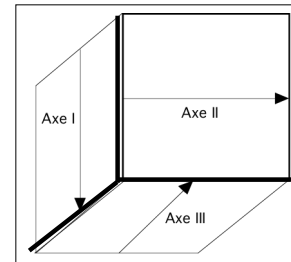
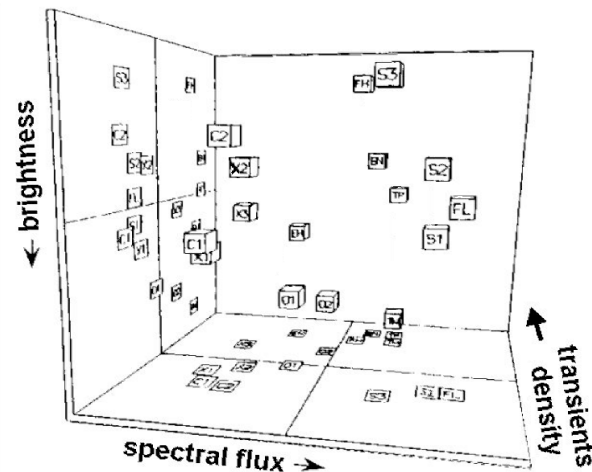
**J. Grey (1977)** recorded, digitized and then analyzed tones from 16 instruments played with equal pitch, loudness, and subjective duration. Listeners then rated the dissimilarities for all pairs of tones.

Grey settled on a three-dimensional structure as capturing the greatest amount of the variation in the data structure while not having so many dimensions as to make the structure difficult to interpret. Physical parameters corresponding to the three dimensions were determined afterwards through analysis of the timbre space.

Grey's timbre space (Grey & Moorer, 1977, p. 1496)

#### Sounds from 16 musical instruments

- O = Oboe
- C = Clarinet
- (1=M**i**, 2=bass)
- X = Saxophone
- (1=f, 2=mf, 3=soprano)
- EH = English Horn
- FH = French Horn
- S = Cello
- (1= *sul tasto*, 2=*normale*, 3=*sul ponticello*)
- TP = Trumpet
- TM = Trombone
- FL = Flute
- BN = Bassoon



**Dimension I (top-bottom)** represents spectral envelope or brightness (brighter sounds at bottom)

Low brightness : french horn and cello *sul tasto* High brightness : oboe, muted trombone (son à qualité stridente, spectre large)

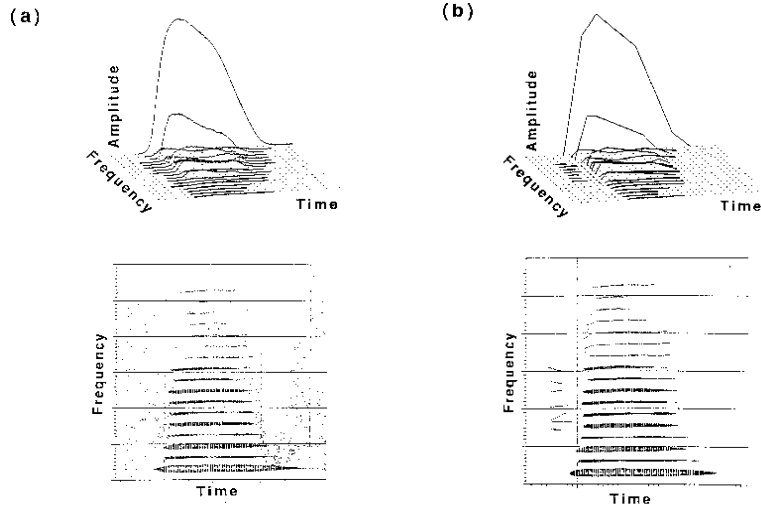
**Dimension II (left-right)** represents spectral flux (greater flux to the right). This dimension is related to a combination of the degree of fluctuation in the spectral envelope and the synchronicity of onsets of the different harmonics (*spectral smoothness*).

High synchronicity and low fluctuation : clarinet, saxophone  
Low synchronicity and high fluctuation : flute, violoncelle

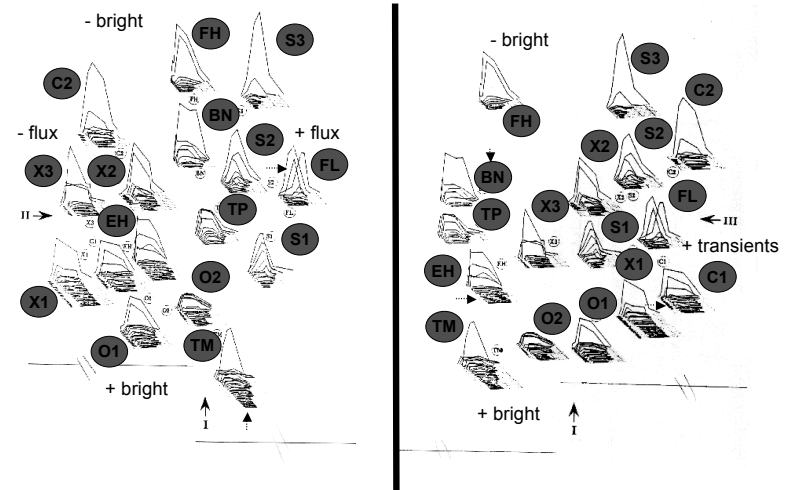
**Dimension III (front-back)** represents degree of presence of attack transients (more transients at the front). It can be called the attack quality.

More transients : strings, flute, single reed instrument (clarinet & saxophone)  
Fewer transients : brass, bassoon, english horn

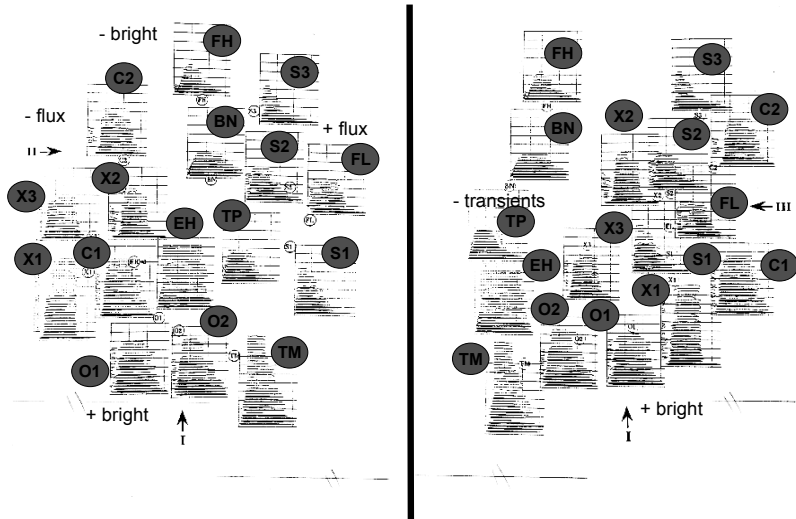
### Representations used for projections



### Bidimensional projections (with time-frequency perspectives)



### Bidimensional projections (with spectrogrammes)



### Conclusion :

Previous research had demonstrated the importance of the energy distribution in spectrum (cf. Bismark). Grey confirms this, since brightness is one the dimensions of the timbre space he determined. He also shows that two temporal features (spectral flux and transient density) can possibly explain grouping of instrumental timbres in families.

Flute and bassoon are interesting cases :

- the attack of a flute (FL) sound is similar to the strings' attack (S1,S2,S3)
- the bassoon (BN) has a profile similar to brass instruments (X,FH,TP).

This would explain the surprising location of these two instruments in space timbre.

### McAdams timbre space (1995)

hrn : horn  
 tpt : trumpet  
 tbn : trombone  
 hrp : harp  
 tpr : trumpar (trumpet/guitar)  
 ols : oboleste (oboe/celesta)  
 vbs : vibraphone  
 sno : striano (bowed string/piano)  
 hcd : harpsichord  
 ehm : English horn  
 bsn : bassoon  
 cnt : clarinet  
 vbn : vibrone (vibraphone/trombone)  
 obc : obochord (oboe/harpsichord)  
 gtr : guitar  
 stg : bowed string  
 pno : piano  
 gnt : guitarnet (guitar/clarinet)

