**Timbre perception**

Kong et al., 2011:

While the temporal envelope based coding strategies used in cochlear implants can support speech recognition, they do not suffice for pitch perception or music appreciation. Timbre is the perceptual attribute that distinguishes two sounds that have the same pitch, loudness and duration. In music, timbre distinguishes instruments playing the same note with the same loudness. Cochlear implant users show poorer recognition (40-60%) of different instruments compared to normal-hearing listeners (90-100%).

This study indicated that the NH timbre space can be represented by three dimensions (3D), with each dimension associated with an acoustic feature. The first timbre dimension was highly correlated with the log-attack time of the stimuli and the second dimension was highly correlated with the spectral centroid

(i.e., the spectral center of gravity) of the stimuli. The third dimension, however, was only weakly correlated with spectral flux (i.e., the fluctuation of the spectrum over time). A more recent study by the same group of researchers (Caclin et al., 2005) investigated the psychophysical nature of timbre dimensions by varying individual acoustic properties (i.e., log-attack time, spectral centroid, and spectral flux or spectral irregularity) of harmonic complexes independently. They concluded that when attack time, spectral centroid, and spectral flux were manipulated independently in the stimuli set, the NH timbre space was best fit with a two-dimensional (2D) model. While spectral flux alone contributes to timbre perception, it was not a robust cue for NH listeners to perceive differences in timbre when all three cues were present in the stimuli set. On the other hand, when attack time, spectral centroid, and spectral irregularity (differences in amplitude between adjacent harmonics) were manipulated independently in the stimulus set, the NH timbre space was best fit with a 3D model with one dimension significantly correlated with the attack time, one dimension correlated with the spectral centroid, and one dimension correlated with the spectral irregularity. Based on these results, timbre can be described as a characteristic or quality of sound that depends on both temporal and spectral aspects of sound, both independently and interdependently.

There is general agreement that the temporal and spectral envelopes are the dominant cues of timbre perception. Cochlear implant users have been found to perform as well as normal hearing controls at discriminating independent temporal and spectral acoustic features.

Results from the current study show that normal hearing listeners use temporal envelope, spectral envelope (the shape of the power spectrum) and spectral fine structure cues (temporal cues meant the changes in amplitude and frequency over time) to perceive differences in musical timbre, consistent with previous research.

the weak or lack of correlation between spectral centroid/spectral irregularity and the second dimension in the CI group data suggests that the spectral cues were less reliable cues for timbre perception in CI listeners compared with their NH counterparts. Indeed, the patterns of results in the CI group data are supported by the findings from individual CI listeners, where only half of the CI listeners showed significant correlation between a perceptual dimension and spectral envelope and spectral fine structure parameters. The other half of the listeners, on the other hand, relied mainly on temporal envelope cues to perceive differences in musical timbre.

McAdams (2013) book:

Timbre is a misleadingly simple and exceedingly vague word encompassing complex auditory attributes and psychological percepts. It includes many percepts not accounted for by pitch, loudness, spatial position, duration or environmental cues like reverb. Timbre is und3erstood as having two characteristics. The first is its perceptual attributes which change in time such as attack sharpness, brightness, nasality, richness and other more discrete or categorical (i.e some characteristic that belongs to a specific instrument. The second is that timbre is an important perceptual cue for recognition, identification and tracking over time of a sound source and thus involved in the absolute categorisation of a sounding object.

Some models of timbre space include individual and class differences as weighting factors on the different dimensions and the set of specificities. For example, some listeners might pay more attention to spectral properties than to temporal ones, while others might not. This variability could reflect either variabilities in sensory processing or in listening and rating strategies.

The most ubiquitous correlates derived from musical instrument sounds include spectral centroid (representing the relative weights of high and low frequencies and corresponding to timbral brightness or nasality: an oboe has a higher spectral centroid than a French horn; see Figure 4), the logarithm of the attack time (distinguishing continuant instruments that are blown or bowed from impulsive instruments that are struck or plucked; see Figure 5), spectral flux (the degree of evolution of the spectral shape over a tone’s duration which is high for brass and lower for single reeds; see Figure 6), and spectral deviation (the degree of jaggedness of the spectral shape, which is high for clarinet and vibraphone and low for trumpet; see Figure 7).

Audio descriptors based on energy envelope, short-term Fourier transform, harmonic sinusoidal components, or the gamma-tone filter bank model of peripheral auditory processing capture temporal spectral, spectrotemporal, and energetic properties of acoustic events. Temporal descriptors include properties such as attack, decay, release, temporal centroid, effective duration, and the frequency and amplitude of modulation on the energy envelope. Spectral shape descriptors include measures of centroid, spread, skewness, kurtosis, slope, crest factor, and jaggedness of the spectral envelope. Spectrotemporal descriptors include spectral flux. Energetic predictors include harmonic energy, noise energy and statistical properties of the energy envelope. There are also descriptors related to periodicity/harmonicity and noisiness. Certain of these descriptors have a single value for a sound event, such as attack time, whereas others represent time-varying quantities, such as the variation of spectral centroid over the duration of a sound event. One problem with a large number of descriptors is that they may be correlated among themselves for a given set of sounds, particularly if they are applied to a limited sound set.

Ask Andrew about that:::::::::::::::::::::::::::>

Most timbre space studies have restricted the pitch and loudness to single values for all of the instrument sounds compared in order to focus listeners’ attention on timbre alone. An important question arises, however, concerning whether the timbral relations revealed for a single pitch and/or a single dynamic level hold at different pitches and dynamic levels and, more importantly for extending this work to real musical contexts, whether they hold for timbres being compared across pitches and dynamic levels.

It is clear that for many instruments the timbre varies as a function of pitch because the spectral, temporal and spectrotemporal properties of the sounds covary with pitch. Listeners are also able to ignore pitch differences within an octave when asked to compare only the timbres of the tones. When the pitch variation is greater than an octave, interactions between the two attributes occur. Marozeau and de Cheveigne´ (2007) varied the brightness of a set of synthesized sounds, while also varying the pitch over a range of 18 semitones. They found that differences in pitch affected timbre relations in two ways: (1) pitch shows up in the timbre space representation as a dimension orthogonal to the timbre dimensions (indicating simply that listeners were no longer ignoring the pitch difference), and (2) pitch differences systematically affect the timbre dimension related to spectral centroid.

Krumhansl and Iverson (1992) found that speeded classifications of pitches and of timbres were symmetrically affected by uncorrelated variation along the other parameter. These results suggest a close relation between timbral brightness and pitch height and perhaps even more temporally fine-grained features related to the coding of periodicity in the auditory system or larger-scale timbral properties related to the energy envelope. This link would be consistent with underlying neural representations that share common attributes, such as tonotopic and periodicity organizations in the brain.

Similarly to pitch, changes in dynamics also produce changes in timbre for a given instrument, particularly, but not exclusively, as concerns spectral properties. Sounds produced with greater playing effort (e.g., fortissimo vs. pianissimo) not only have greater energy at the frequencies present in the softer sound, but the spectrum spreads toward higher frequencies, creating a higher spectral centroid, a greater spectral spread, and a lower spectral slope. No studies to date of which we are aware have examined the effect of change in dynamic level on timbre perception, but some work has looked at the role of timbre in the perception of dynamic level independently of the physical level of the signal.

Regarding the second approach to timbre, meaning its role in identification of sound source like musical instruments, one reasonable hypothesis is that the sensory dimensions that compose timbre serve as indicators used in the categorisation, recognition, and identification of sound events and sound sources.

Honestly not gonna write all this stuff but basically the section on timbre as a vehicle for source identity is very useful.

Timbre perception is at the heart of orchestration, a realm of musical practice that has received relatively little experimental study or even music-theoretic treatment for that matter. Instrumental combinations can give rise to new timbres if the sounds are perceived as blended. Timbral differences can also both create the auditory streaming of similar timbres and the segregation of dissimilar timbres, as well as induce segmentations of sequences when timbral discontinuities occur. Listeners can perceive intervals between timbres as similar when they are transposed to a different part of timbre space, even though such relations have not been used explicitly in music composition. Timbre can play a role in creating and releasing musical tension. And finally, there is some evidence that listeners can learn statistical regularities in timbre sequences, opening up the possibility of developing timbre-based grammars in music.

Jiang et al., 2020:

McAdams et al. studied the common dimensions of timbre spaces with synthetic sounds used as experimental materials, establishing a relationship between the dimensions of a space and the corresponding acoustic parameters [52]. Martens et al. used guitar timbre to study the differences in timbre spaces constructed under different language backgrounds [53,54]. Zacharakis and Pastiadis conducted a subjective evaluation and analysis using 16 Western musical instruments, proposing a luminance–texture–mass (LTM) model for semantic evaluation. In this process, six semantic scales were analyzed using principal component analysis (PCA) and multidimensional scaling (MDS) to produce two different timbre spaces [55]. Simurra and Queiroz used a set of 33 orchestral music excerpts that were subjectively rated using quantitative scales based on 13 pairs of opposing verbal attributes. Factor analysis was included to identify major perceptual categories associated with tactile and visual properties, such as mass, brightness, color, and scattering.

V complicated; they created a 3d timbre space.

Town & Bizley, 2013:

Timbre is operationally defines as the attribute that distinguishes sounds of equal pitch, loudness, location and duration. Functionally, timbre is a key determinant of sound identity and plays a pivotal role in speech as it is the principal determinant of phonetic identity. Despite its importance timbre remains one of the least studies and perhaps most challenging features of sound to understand.

Speech signals contain a wide variety of acoustic cues from which sound timbre may be derived and our perception of any one segment of speech may be influenced by the context in which it occurs. At the phonetic level, timbre plays a crucial role in determining the identity of vowels and consonants.

The underlying features of sound that contribute to differences in perceived timbre can be investigated by comparing the acoustic properties of instruments. Such comparisons indicate that, like vowels and consonants, both spectral and temporal features of sound determine timbre. In the spectral domain, most musical instruments emit harmonic resonances, that is, they produce energy at integer multiples of a fundamental frequency (the harmonics are evident as the horizontal bands in the spectrograms in Figure 1). Such harmonics resemble those introduced in speech by the vibration of the vocal chords. As with vowels, the distribution of energy across different harmonics is one of the key differences between different musical instruments. For example, the piano (Figure 1D) has sustained energy only at the fundamental while the violin and accordion (Figure 1E) has energy distributed over many harmonics, and the oboe contains most of its energy in the first five harmonics (Figure 1F).

Some instruments, such as the clarinet, have energy only in the odd harmonics, whereas notes played by the trombone only have energy at the first and second harmonic (Campbell and Greated, 1994). The second key determinant of the timbre of a musical instrument is its temporal characteristics, or what musicians call the “nature of attack”. This is especially the case for plucked string instruments like the harp, or piano whose notes contain little or no steady state sound at all (Figure 1C; Campbell and Greated, 1994). In this case, the shape of the amplitude envelope at the beginning of the sound will be key to the perceived tone quality.

The acoustic basis of musical timbre has also been studied using multidimensional scaling (MDS) techniques (Plomp and Steeneken, 1969; Miller and Carterette, 1975; Grey, 1977; Wessel, 1979; McAdams, 1999; McAdams and Giordano, 2009). Simply put, MDS aims to determine the components and underlying structure of a data space from a series of distance measures. Here, the space of interest is the perceptual representation of musical timbre and the distance measures are dissimilarity judgments of listeners to instrument sounds. After constructing a timbre space using MDS, it is possible to relate different perceptual dimensions back to the acoustic features of the instrument sounds. For example, Grey (1977) found that similarity judgments of synthesized instrument sounds could be mapped into a three-dimensional space in which dimensions were strongly correlated with spectral shape, the presence of low-amplitude high-frequency energy in the attack segment of sounds and the combination of spectral fluctuation with synchronous high frequency transients. Numerous MDS studies since have replicated the finding that the spectral shape of instrument sounds (or related statistics such as spectral centroid) and attack time are important components of timbre spaces (McAdams et al., 1995; Lakatos, 2000; Caclin et al., 2005; Burgoyne and McAdams, 2008).

Timbre not only enables musical instrument identification, but also enables listeners to estimate the scale of an instrument. In addition to listeners being able to recognize the family of an instrument sound, even when that sound was modified in pitch and scale beyond the range normally experienced, listeners could accurately assess the scale of a given instrument (van Dinther and Patterson, 2006).

Finally, timbre also plays a role in the perception of environmental sounds such as impacted bars or plates, that is, sounds produced when a bar or plate is struck. Human listeners are able to classify the material, hollowness and, to a lesser extent, size and shape of such bars or plates from the impacted sound alone (Lakatos et al., 1997; Kunkler-Peck and Turvey, 2000; Lufti, 2001; Tucker and Brown, 2003; Lufti, 2007). The classification of a bar or plate’s material may depend on its damping properties—the extent to which oscillations in the plate or bar are reduced in amplitude over time after being struck. For example metal and glass plates differ in the extent to which they are subject to thermo- and viscoelastic damping (Chaigne and Lambourg, 2001). When the sounds of impacted plates are artificially damped by suspension in water, listener’s judgments of material, shape and size become less reliable (Tucker and Brown, 2003; although see also Giordano and McAdams, 2006).