1pPPb7. Acoustic correlates of auditory object perception: Speakers, musical timbres, and environmental sounds. Mattson Ogg and L. Robert Slevc (Neurosci. and Cognit. Sci., Univ. of Maryland, College Park, Biology/Psych. Bldg. Rm. 3150, 4094 Campus Dr., College Park, MD 20742, mogg@umd.edu)

Human listeners must identify and orient themselves to auditory objects in their environment. What acoustic features support a listener's ability to differentiate the variety of sound sources they might encounter? Typical studies of auditory object perception obtain dissimilarity ratings between pairs of objects, often within a single category of sound. However, such an approach precludes an understanding of general acoustic features that might be used to differentiate sounds across categories. The present experiment takes a broader approach to the analysis of dissimilarity ratings by leveraging the acoustic variability within and between different sound categories as characterized by a large, diverse set of 36 sound tokens (12 speech utterances from different speakers, 12 instrument timbres, and 12 everyday objects from a typical human environment). We analyze multidimensional scaling results as well as models of trial- level dissimilarity ratings as a function of different acoustic representations including spectral, temporal and noise features as well as modulation power spectra and cochlear spectrograms. In addition to previously noted differences in spectral and temporal envelopes, results indicate that listener's dissimilarity ratings are also related to spectral variability and noise, particularly in differentiating sounds between categories. Dissimilarity ratings also appear to closely parallel sound identification performance.

1pPPb8. A pitch encoding model based on the intrinsic oscillation circuit: Effects of the time constant and input stimulus types. Minoru Tsuzaki (Kyoto City Univ. of Arts, 13-6 Kutsukake-cho, Oe, Nishikyo-ku, Kyoto 610-1197, Japan, minoru.tsuzaki@kcua.ac.jp) and Katuhiro Maki (Faculty of Human Informatics, Aichi Shukutoku Univ., Aichi, Japan)

Many temporal models for pitch perception have adopted a configuration of delay-lines and coincidence detectors after the cochlear filtering. Autocorrelation functions are a usual way of its implementation. However, a series of experiments by the authors' group have revealed that the perceived pitch would shift upwards by the effect of aging. Because the autocorrelation simply represents the time intervals statistics in the physical domain, the aging cannot affect this statistics. Therefore, a further pitch encoding process where the physical (physiological) temporal intervals are mapped against any internal reference in the brain. We propose a model comprised of bank of self-oscillatory circuits and the coincidence detectors. The periods of oscillations are intrinsic characteristics of the neural circuit. The proposed model could pick up the fundamental periods of a various types of stimuli, i.e., pure tones, missing fundamentals and iterated ripple noises. To check an effect of aging, the changes in the time constant of the temporal waveform of the oscillation was investigated. Although the shortest sensitive period increased due to the longer time constant, the age-induced pitch shift could not be predicted. It is necessary to assume a systematic lengthening of the period of each oscillation to predict the pitch shift.

1pPPb9. Speaker-dependent low-level acoustic feature extraction for emotion recognition. Tejal Udhan and Shonda Bernadin (Elec. Eng., Florida State Univ., 2525 Pottsdamer St., Tallahassee, FL 32310, tu13b@my.fsu.edu)

In this paper, accuracy for emotion recognition using low- level acoustic features is investigated. The aim of any speech emotion recognition system is to extract acoustic features that are representative of the emotional state of the speaker. Frequency formants, intensity, and pitch are the low-level features proposed for characterizing four different emotions, anger, happy, sadness, and neutral, using acoustic data. Low- level features describe the acoustic, prosodic, and spectral properties of the speech signal and limit the complexity of emotion recognition systems. An algorithm is designed for characterizing each emotion using the acoustic features. It has been proven that various aspects of a speaker's physical and emotional state can be identified by speech alone. However, the accuracy of such analyses has not been optimized due to acoustic variabilities such as length and complexity of human speech utterance, gender, speaking styles and speech rate. It has also

been found that speaker-dependent systems are more accurate in emotion recognition than that of speaker-independent systems. Since speech emotion recognition is relatively a newer field, the set of most powerful features which can distinguish different emotions is not defined; hence, examining the accuracy of emotion recognition using selected acoustical features is an important task.

1pPPb10. Predicting timbral and perceptual characteristics of orchestral instrument combinations. Aurelien Antoine and Eduardo Miranda (Interdisciplinary Ctr. for Comput. Music Res. (ICCMR), Univ. of Plymouth, Plymouth University, The House Bldg. - Rm. 304, Plymouth PL4 8AA, United Kingdom, aurelien.antoine@postgrad.plymouth.ac.uk)

Orchestration is a compositional practice that consists of writing for several instruments. This process often involves harnessing each instrument's sound to create sonic textures that could not be achieved with a single instrument. These sound fusions are usually sought by composers to express specific perceptual effects. However, the number of potential combinations is significant. Testing and analyzing all combinations to identify the ones matching the desired perceptual effects is logistically and computationally complex. Using supervised learning methods to create regression and classification models, it is possible to predict specific timbral and perceptual characteristics from information about a combination of different orchestral instruments. Such developments would provide methods to estimate the perception of instrument timbre fusions directly from abstract information. Similar methods could potentially be applied to other types of sources and predict specific perceptual characteristics without the need to perform an acoustical and psychoacoustical analysis on every audio sources.

1pPPb11. Informational masking of calm speech targets by emotional speech maskers. Shae D. Morgan (Commun. Sci. and Disord., Univ. of Utah, 390 South 1530 East, Ste. 1201, Salt Lake City, UT 84112, shae.morgan@utah.edu)

Recent research suggests that some types of emotional speech are more intelligible than others when presented in background noise (Dupuis & Pichora-Fuller, 2015). Yet, emotional speech can often be present as the background noise itself. Attentional mechanisms and auditory stream segregation capabilities likely impact target word recognition in these situations with emotional speech maskers. The present study examined calm speech target sentences in masking background combinations of four emotions (calm, sad, happy, and angry) and two masker types (2-talker babble and speech-shaped noise). The emotion categories differ from one another on perceptual dimensions of activation (low to high) and pleasantness (unpleasant to pleasant). Speech-shaped noise maskers were spectrally and temporally similar to the 2-talker babble maskers. Performance was compared between the two masker type conditions to quantify the amount of informational masking induced by different emotional speech maskers. Furthermore, the number of target-masker confusions were identified for each masker emotion and subsequently analyzed. Results showed lowest target recognition with calm speech maskers, consistent with auditory stream segregation framework; however, confusions and informational masking were high for angry and happy speech maskers, which suggests that auditory attention to high-arousal stimuli may have influenced performance.

1pPPb12. Mechanical sound similarity and categorization. Annamarie Rosen, Nirmal Kumar Srinivasan (Audiol., Speech-Lang. Pathol., and Deaf Studies, Towson Univ., 8000 York Rd., Towson, MD 21252, nsrinivasan@towson.edu), Jeremy Gaston, and Brandon Perelman (Human Res. and Eng. Directorate, U.S. Army Res. Lab., Adelphi, MD)

Recent studies suggest that perceptual similarity among sounds can be used to predict accuracy in perceptual tasks (Dickerson & Gaston, 2014; Gaston, *et al*., 2017). We operationalize this similarity as distance in a 2D similarity space generated using multidimensional scaling (MDS) on pairwise similarity stimulus ratings. We tested listeners' ability to categorize mechanical environmental sounds, and modeled the relationship between stimulus categorization and similarity. Stimuli were 18 mechanical sounds selected hierarchically from three broad categories (vehicles, power tools, and household appliances), and three subcategories. Thirty-five normal