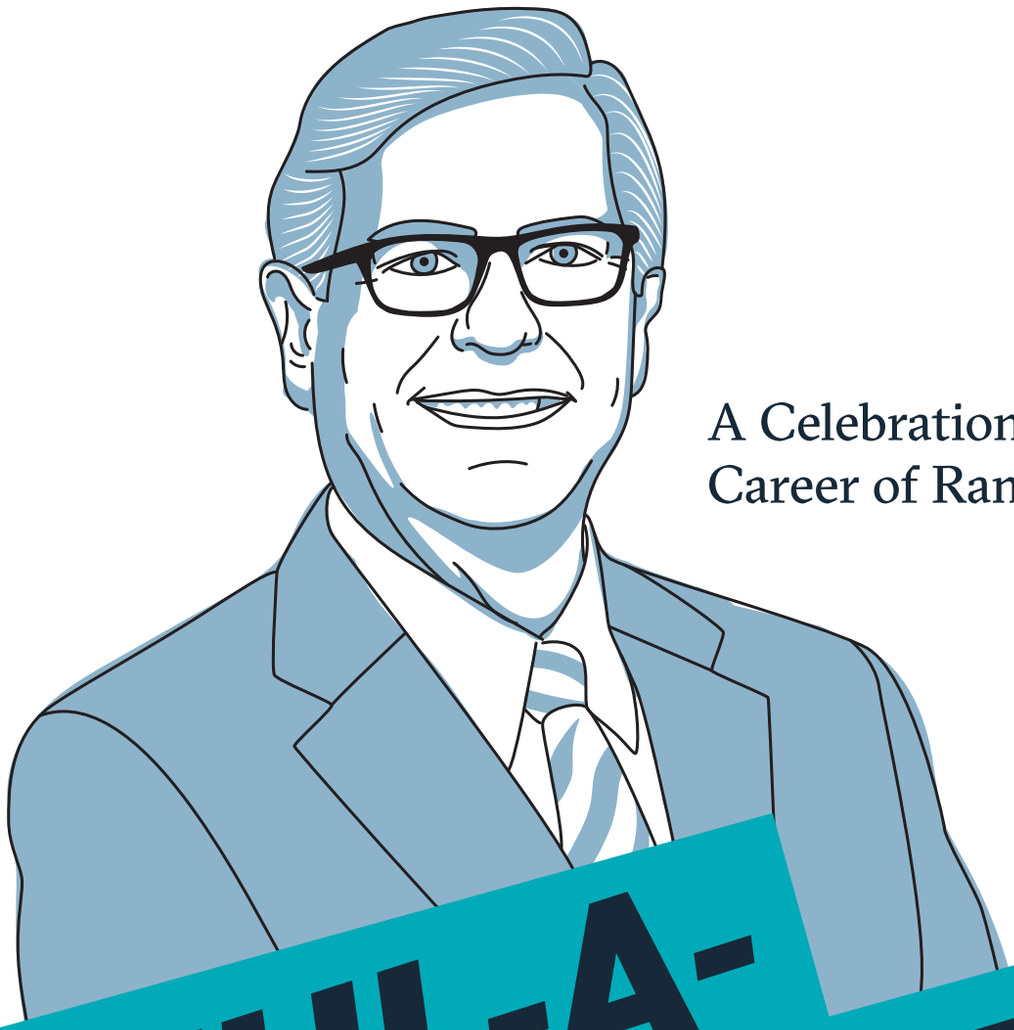

A Celebration of the
Career of Randy Diehl

DIEHL-A-PALOOZA

SPONSORED BY:

The University of Texas at Austin
Center for Perceptual Systems
College of Liberal Arts

April 20-21, 2018

DIEHL-A-PALOOZA PROGRAM

FRIDAY, APRIL 20	
8:30 am	COFFEE, PASTRIES, & SIGN IN
9:00 am	Welcome and Announcements
9:15 am	Lori L. Holt (Carnegie Mellon University) *Cues Worth Weighting For*
9:30 am	Bharath Chandrasekaran (UT Austin) *Neural Systems in Auditory and Speech Categorization*
10:00 am	Amanda Rysling (UC Santa Cruz) & John Kingston (UMass Amherst) *Order Matters*
10:30 am	TOAST / ROAST / STORIES
10:45 am	COFFEE BREAK
11:00 am	President Gregory L. Fenves
11:15 am	Melissa A. Redford (University of Oregon) *Measuring Production with Perception*
11:45 am	Andrew Lotto (University of Florida) *The Scientific Virtues of Randy Diehl and the Decline of Theoretical Speech Science*
12:00 pm	LUNCH
1:45 pm	Patrick Wong (Chinese University of Hong Kong) *Variability in Speech Learning*
2:15 pm	Wilson Geisler (UT Austin) *Decision-Variable Correlation: How Correlated Are Your Decisions with Randy's?*
2:30 pm	TOAST / ROAST / STORIES
3:00 pm	COFFEE BREAK
3:30 pm	David W. Gow, Jr. (Massachusetts General Hospital/Harvard Medical School/Salem State University) *Phonetic Contrast and Lexical Similarity: Complementary Influences on Phonology from the Neural Mechanisms That Support Speech Perception and Spoken Word Recognition*

4:00 pm	Christian Stilp (University of Louisville) *Contrast As a Fundamental Premise Underlying Speech Perception*
4:30 pm	Jessica S. F. Hay (University of Tennessee, Knoxville) *(Perceptual) Learning Begets (Language) Learning: A Speech Perception Researcher Diehls in Development*
5:00 pm	TOAST / ROAST / STORIES AND CLOSING REMARKS
6:00 pm	RECEPTION, RANDY'S RESPONSES, AND GROUP PHOTO
SATURDAY, APRIL 21	
9:00 am	COFFEE AND PASTRIES
9:30 am	Tyler K. Perrachione (Boston University) *Social Consequences of Phonetic Expertise: The Language-familiarity Effect in Talker Identification*
10:00 am	Michelle Molis (NCRAR, Oregon Health & Science University) *How Phonology Can Inform Psychoacoustics and Hearing Science*
10:30 am	COFFEE BREAK
11:00 am	Francisco Lacerda (Stockholm University) *Peripheral Auditory-Nerve Adaptation and the Discrimination of Speech Sounds – Revisiting 30 Years Later*
11:30 am	Keith R. Kluender (Purdue University) *Virtues of (Co)variance for Perceptual Learning*
11:45 am	TOAST / ROAST / STORIES
12:15 pm	LUNCH
2:00 pm	Adrian Garcia-Sierra (University of Connecticut) *The Context That Establishes the Boundary: An ERP Study Assessing the Perception of Stop Consonants in Bilinguals and Monolinguals in Two Different Phonetic Contexts*
2:30 pm	Megan Crowhurst (UT Austin) *Rhythm and Cues: Creak, Duration, and Extensions of the Iambic-Trochaic Law*
3:00 pm	Beverly A. Wright (Northwestern University) *Induction and Prevention of Speech Learning*
3:30 pm	COFFEE BREAK

4:00 pm	Laurel H. Carney & Joyce M. McDonough (University of Rochester) *Nonlinear Auditory Models Yield New Insights into Representations of Vowels*
4:30 pm	Bob McMurray, K.E. Schreiber, J. Klein, & M.E. Galle (University of Iowa) *Who Knows Where the Time Went? Temporal Integration of Acoustic Input During Spoken Word Recognition is Not Always Linear*
5:00 pm	Rajka Smiljanic (UT Austin) *Intelligibility Variation: Talkers, Listeners, and Signals*
5:30 pm	CLOSING REMARKS
7:00 pm	SPEAKERS' DINNER

Organizers: John Kingston, Lori Holt, Andrew Lotto

Abstracts

Cues Worth Weighting For
Lori Holt, Department of Psychology, Carnegie Mellon University

Early research in speech perception focused on discovering the acoustic dimensions that would invariantly convey phonemes. From this starting point, the theoretical and empirical contributions originating from Randy's lab have seeded the intellectual foundations of contemporary research. This work takes acoustic variability seriously as a 'feature' rather than a 'bug.' It appreciates the contributions of both long- and short-term learning on how acoustic input is mapped to phonetic categories. And, it acknowledges the importance of general perceptual constraints and a neurobiology shared with other auditory processing. I will review research from my lab and those of others in a tour of how Randy's research contributions have shaped contemporary understanding of the mapping from speech acoustics to phonemes.

Neural Systems in Auditory and Speech Categorization
Bharath Chandrasekaran, Departments of Communication Sciences & Disorders, Psychology, Linguistics; Institutes of Mental Health Research and Neuroscience, University of Texas, Austin

Speech sounds are multidimensional, acoustically variable, and temporally ephemeral. Despite the enormous computational challenge, native speech perception is rapid and automatic. Over the last decade, we have made substantial progress in our understanding the cortical mechanisms underlying mapping of speech onto meaning (ventral pathway) and articulation/location (dorsal p[athway). The dorsal and ventral streams are useful points of reference to understand the processing of native speech signals. The primary goal of this talk is to elucidate mechanisms underlying how novel non-speech and speech categories are acquired and represented in the mature brain through the time course of learning. The talk will discuss a novel theoretical framework, the dual-learning systems (DLS) model which characterize the neurobiology of two complementary corticostriatal auditory streams: a sound- to-rule mapping stream and a sound-to-reward mapping stream that are critical in feedback-dependent learning of novel auditory categories. We posit that temporal lobe circuits are trained by the dual corticostriatal circuits to incorporate dimensional rules (via the sound-to-rule stream), multidimensional integration (via the sound-to-reward stream), ultimately leading to automatic, and abstract category representations. I will discuss several experiments that will systematically test the DLS model: First, I will characterize differences in the neural circuitry underlying the acquisition of two non-speech auditory categories that are structured to be optimally learned by either the sound-to-rule or the sound-to- reward streams. Second, I will examine the role of these dual streams in mediating the acquisition of new (non-native) speech categories across different levels of expertise (from novice to native) using multimodal neuroimaging methods. Finally, I will discuss several experiments that leverage the neurobiology of corticostriatal auditory streams to design optimal training paradigms that reduce inter- individual differences in auditory category learning. Overall our systems neuroscience approach will delineate the role of multiple, functionally distinct corticostriatal loops in audition. The results from these experiments will provide important scaffolding for the understanding of the impact of striatal dysfunction in individuals with auditory and speech processing disorders.

Order Matters

Amanda Rysling, Linguistics Department, University of California, Santa Cruz
John Kingston, Linguistics Department, University of Massachusetts, Amherst

We report a novel observation: the order of contexts and targets determines how the contexts influence perception of targets in speech sound categorization studies. When both contexts and targets vary in the distribution of energy in their spectra, spectral contrast is only found when contexts precede targets. When contexts instead follow targets, assimilation is found: listeners judge the ambiguous steps of a target continuum to be spectrally similar to following contexts. Assimilatory effects are modulated by the degree of acoustic continuity between targets and contexts. These results are consistent with stimulus-specific adaptation as the underlying mechanism for spectral contrast. Assimilation instead arises earlier in perceptual processing and is a product of listeners continuing to attribute the signal's acoustic properties to structure in the perceptual representation that they've already built until a sufficient acoustic discontinuity leads them to start building new structure. In an acoustically continuous transition between target and following context, the acoustic properties of the context would thus be attributed to the target, producing assimilation rather than contrast. This finding pushes past the debate between gesturalist and auditorist accounts of speech perception and forces reanalysis of sound changes that have been argued to be byproducts of hypocorrection.

Measuring Production with Perception

Melissa A. Redford, University of Oregon

I became an honorary member of Randy's lab in the mid 1990s when we conducted a perceptual confusion study to test the hypothesis that syllable-final consonants are typologically disfavored relative to syllable-initial consonants because listeners are less able to recover the former than the latter (Redford & Diehl, 1999). The results confirmed the hypothesis. Follow-up acoustic analyses strongly suggested that the reason for the perceptual difference lay in production. Since that study, I have often used perception to inform my work on production because listener ratings or judgments can provide a single, synthetic measure of the complex, multivariate information present in speech. Recently, I have built on this approach to measure coarticulation using audiovisual speech perception. Questions about coarticulatory scope are tested in a gating paradigm: perceivers are presented with information up to experimentally-defined critical junctures in the speech stream and then asked to predict an upcoming or preceding sound. Coarticulatory strength is measured as the extent to which perceivers are correct in their predictions.

The Scientific Virtues of Randy Diehl and the Decline of Theoretical Speech Science

Andrew J. Lotto, Department of Speech, Language, and Hearing Science, University of Florida

I will present a short overview of some of the contributions of Randy Diehl that highlight his commitment to theoretically-driven empirically-constrained science. I will also contrast his approach to distributional or "statistical" learning and adaptation in speech perception with current approaches that often fail to be constrained enough to provide meaningful a priori predictions.

Variability in Speech Learning

Patrick Wong, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong

The vast majority of language users speak at least one spoken language. Decades of research has been devoted to answering many questions about spoken language, from questions related to its basic structure and units of abstraction, psychological realization, to biological foundation. One intriguing aspect of spoken

language that drives this large body of research and the accompanied theoretical advancement is variability. There are at least two sources of variability. First, the acoustic signal produced by one talker can have a very different acoustic surface realization compared to the signal of the same phonetic category produced by another talker. Second, listeners differ in how they construct a speech category even when the speech signal is identical. In this presentation, I will highlight some of the studies from my research group that has contributed to the body of work that examines variability in speech. From infancy to older adulthood and across typical and atypical populations, it becomes increasingly clear what the neurological underpinnings of variability may be and how an understanding of this variability may allow for constructing models of systems neuroscience of language that capture the range of language performance we observe. I argue that an emphasis on investigating variability also has tremendous potential for translational linguistics that has direct implications for clinical and pedagogical practices.

Decision-Variable Correlation: How Correlated Are Your Decisions with Randy's?
Wilson S. Geisler, Department of Psychology, University of Texas, Austin

A straightforward extension of the signal detection theory (SDT) framework is described and demonstrated for the two-alternative identification task. The extended framework assumes that the subject and an arbitrary model (or two subjects, or the same subject on two occasions) are performing the same task with the same stimuli, and that on each trial they both compute (in effect) values of a decision variable. Thus, their joint performance is described by six fundamental quantities: two levels of intrinsic discriminability ("d-prime"), two values of decision criterion, and two decision-variable correlations, one for each of the two categories of stimuli in the task. Decision-variable correlations provide increased power for testing models and for characterizing individual differences, and do not require a special experimental design (e.g., they can be computed from existing data). The extended framework was developed for analyzing trial-by-trial performance in vision experiments with natural stimuli, but it should be widely applicable in behavioral and neurophysiological studies of perception and cognition.

Phonetic Contrast and Lexical Similarity: Complementary Influences on Phonology from the Neural Mechanisms That Support Speech Perception and Spoken Word Recognition
David W. Gow Jr., Massachusetts General Hospital/Harvard Medical School/Salem State University

Phonology is shaped by the ways that neural mechanisms meet the demands of robustly recognizing an inherently noisy speech signal. Robust speech perception needs contrast. Randy Diehl's powerful work on auditory enhancement shows how the general operating principles of the auditory system define and enhance perceptual contrast and ultimately select speech sound inventories. Here, I would like to argue that the mechanisms that enable spoken word recognition introduce a balancing pressure towards similarity that also contributes to the structure of phonological systems. Interactivity is a fundamental property of cortical processing. In spoken word recognition, interactivity introduces a variety of contextual constraints that minimize perceptual uncertainty and coordinate parallel analyses of higher level linguistic and general cognitive analysis. By biasing the interpretation of speech sounds to fit existing lexical categories, interactive processes produce phonotactic preferences in perception and production that reflect the structure of the lexicon. These biases are self-reinforcing, as lexically-induced biases ultimately shape the structure of the lexicon itself.

Contrast As a Fundamental Premise Underlying Speech Perception
Christian Stilp, Department of Psychological and Brain Sciences, University of Louisville

Human speech perception is a remarkable accomplishment considering the constant challenges posed by acoustic variability, uncertainty, background noise, and other impediments to successful communication. The uniqueness of human speech perception begged for accounts and mechanisms that could explain this proficiency. Early accounts proposed mechanisms that were uniquely human and tied directly to speech production. However, parallel patterns of perception using speech and nonspeech stimuli across humans and nonhuman animals implied that the underlying auditory processing must be more rudimentary in nature. Diehl and colleagues have provided several demonstrations that contrast, or the perceptual emphasis of acoustic differences between successive stimuli, has tremendous explanatory power for understanding speech perception. Here, I will extend this premise by showing that spectral contrast effects are a pervasive influence in perception of speech and nonspeech sounds in healthy and impaired hearing alike. The robustness of these effects provides a strong foundation for studying perceptual accommodation of the listening environment, natural signal statistics, talker characteristics, and more.

(Perceptual) Learning Begets (Language) Learning: A Speech Perception Researcher Diehls in Development
Jessica S. F. Hay, Department of Psychology, University of Tennessee, Knoxville

At its foundation acquiring language involves perceptual learning. Infants are bombarded with perceptual information that they need to learn to interpret in order to make sense of the world around them. However, interpreting perceptual information is not an easy task, given that our sensory environment is exceedingly complex and variable, often lacking explicitly accessible structure. Fortunately, infants are able to use their perceptual systems to begin chipping away at the complexity of the world around them by tracking statistics in their environment to learn some of the most fundamental features of their native language(s), such as what sounds are relevant, how sounds go together, and how sound sequences map onto visual referents, and so on. As a former graduate student of Randy Diehl's, I have applied his influential perceptual learning approach to the study of early language acquisition. In my work, I show that not only do infants' perceptual systems filter the incoming perceptual information, but also that infants' perceptual systems are themselves altered through experience. Specifically, in a series of studies with young toddlers, I provide evidence that learning about the sounds and sound co-occurrence patterns in language drives subsequent mapping of sounds to meaning. In this way, learning the perceptual features of the native language impacts subsequent learning in a non-trivial way.

Social Consequences of Phonetic Expertise: The Language-familiarity Effect in Talker Identification
Tyler K. Perrachione, Department of Speech, Language, and Hearing Sciences, Boston University

Listeners' phonetic expertise in their native language not only helps them perceive speech, it also helps them identify talkers by the sound of their voice. Compared to talkers speaking listeners' native language, talkers speaking an unfamiliar or foreign language are disproportionately more difficulty to tell apart – a phenomenon known as the language-familiarity effect in talker identification. In this talk, I will explore what we know about the contribution of phonetic knowledge to talker identification, with an emphasis on how phonetic expertise allows us to better identify voices speaking our native language, as well as what information listeners need in order to improve their ability to identify foreign-language voices. Research on the language-familiarity effect suggests a parallel and hierarchical model of voice identification: Listeners draw on information from vocal source acoustics, articulatory phonetics, and higher-level linguistic knowledge for optimally accurate talker identification.

How Phonology Can Inform Psychoacoustics and Hearing Science

Michelle R. Molis, VA RR&D National Center for Rehabilitative Auditory Research (NCRAR), Department of Otolaryngology, Oregon Health & Science University

Auditory psychophysicists rarely consider articulatory constraints or phonological systems. More often, their objects of investigation are the sensory and perceptual thresholds for simplified acoustic characterizations of speech or for generalized phonetic realizations. As demonstrated by Diehl and colleagues, models of preferred vowel systems based on predictions of auditory distinctiveness can be improved by taking into account the acoustic properties of the environment and more realistic auditory representations. Hearing impairment presents a unique listening context that will alter the influence of environmental noise and limit access to sufficient auditory distinctiveness in speech. Impaired auditory processing will render some phonetic cues inaudible and will distort others through loss of spectral or temporal resolution. Adult listeners with acquired hearing loss must combine degraded access to auditory cues with their previous language experience to map altered auditory representations onto preexisting phonologic systems. Consequently, inclusion of phonetic and phonologic considerations may improve the prediction of perceptual confusions and inform future auditory rehabilitation strategies.

Peripheral Auditory-Nerve Adaptation and the Discrimination of Speech Sounds – Revisiting 30 Years Later

Francisco Lacerda, Department of Linguistics, Stockholm University

In my PhD thesis (Lacerda, 1987) I used behavioral methods to investigate the potential perceptual impact that the peripheral auditory-nerve fibers' adaptation might have on the discrimination of speech sounds. The theme was inspired by the neurophysiological work of Bertrand Delgutte and Nelson Kiang. Delgutte recordings from cats' auditory nerve fibers had revealed dramatic changes in the fibers dynamic range, just after about 20 ms after the onset of an abrupt stimulus, and that the dynamic range was reduced even in responses to stimuli with gradual onsets. Delgutte and Kiang studied the discharge patterns associated with the auditory nerve fibers representation of different types of speech sounds with different types of onsets, like fricatives and affricates (Delgutte, 1980, 1984; Delgutte & Kiang, 1984a, 1984b, 1984c, 1984d) and consistently found that the dynamic range of the representations was initially larger for sounds with abrupt onset than for those with gradual onsets. The thesis attempted to address the perceptual consequences of that rapid auditory-nerve fiber adaptation in the context of the perception of speech sounds. It studied human perception and tried to contribute with behavioral data to the then ongoing debate on the nature of categorical perception phenomena (Bertoncini, Bijeljac-Babic, Blumstein, & Mehler, 1987; Blumstein & Stevens, 1981; Diehl & Kluender, 1987; Kluender, Diehl, & Killeen, 1987; Molfese, 1978; Pisoni, 1977; Pisoni & McNabb, 1974; Stevens, 1972). The conclusion was that certain patterns of discrimination and categorization of speech sounds could indeed be attributed to potential differences in their underlying discharge patterns at the peripheral auditory-nerve fibers' level. The present contribution is a reassessment of the thesis behavioral data, now with the potential support of additional ABR and ERP data and the question is: Will the thesis conclusions hold and articulate with new EEG-data 30 years later?

Virtues of (Co)variance for Perceptual Learning

Keith R. Kluender, Department of Speech, Language, and Hearing Sciences, Purdue University

Because the world has structure, events in the sensory environment are generally predictable, and most of the energy impinging upon sensory transducers is redundant. Consequently, efficient sensory systems exploit predictability to optimize sensitivity to less predictable inputs that are, by definition, more informative. Growing evidence reveals sensitivity to co-occurrence of stimulus attributes within stimuli. Speech signals

are notoriously redundant, and both simulations and experiments reveal that auditory perception rapidly reorganizes to efficiently capture this redundancy (covariance) among stimulus attributes. Listeners' ability to distinguish sounds from one another is driven primarily by the extent to which they are consistent or inconsistent with patterns of covariation among stimulus attributes. These findings have implications for perceptual learning most broadly, for experience-dependent auditory development and plasticity, for learning of speech sound contrasts, and for learning to talk.

The Context That Establishes the Boundary: An ERP Study Assessing the Perception of Stop Consonants in Bilinguals and Monolinguals in Two Different Phonetic Contexts

Adrian Garcia-Sierra, Department of Speech, Language, and Hearing Sciences, University of Connecticut

Previous research has shown that bilingual speakers of Spanish and English adjust their perception of stop consonants to match the language they are hearing in a given moment. In the present investigation, we report brain responses associated with the processing of stop consonants occurring in two different phonetic contexts. For this purpose, Event Related Potentials (ERPs) were recorded in 20 monolingual speakers of English and 20 bilingual speakers of English and Spanish. Participants' brain responses were collected using a multi-deviant paradigm where 3 target sounds occurred in two distinct contexts. In the Spanish phonetic context condition, the target sounds were presented along with 6 background sounds with negative VOT (Spanish like consonants). In the English phonetic context condition, the 3 target sounds were presented along with 6 background sounds with positive VOT (English like consonants). Each background sound followed a target sound equally often, but keeping stimulus presentation otherwise random. We explored if the perception of target sounds changed between phonetic contexts. The results showed that bilinguals, but not monolinguals, perceived target sounds accordingly with the phonetic contexts regardless of the acoustic separation between background and deviant sounds. The brain mechanisms associated with bilinguals' adjustment in perception are discussed.

Rhythm and Cues: Creak, Duration, and Extensions of the Iambic-Trochaic Law

Megan Crowhurst, Department of Linguistics, University of Texas, Austin

Recently published reports suggest that humans' use of creaky voicing and duration to segment rhythmically patterned sequences into smaller, recurrent units can be influenced by linguistic experience. Our ability to group rhythmic sequences is shared with nonhuman mammals whose grouping biases can also be shaped by experience. For example, Long-Evans rats, a species with no demonstrated language capacity, have been trained to respond selectively to either trochaic long-short or iambic short-long tone groupings to acquire food. It is therefore interesting to consider experiential factors other than language background which might influence humans' biases when segmenting rhythmic sequences in which nonlinguistic analogues of creaky voicing, in particular, are varied. This presentation reports the outcome of a study which explored the influence of the extent and type of musical experience on human listeners' segmentation behaviour. Three groups of English speakers were tested: musically untrained listeners, highly trained jazz/funk musicians, and highly trained musicians without a jazz/funk focus were tasked with grouping sequences of trumpet notes in which duration and a creak-like quality ("growl") were rhythmically varied. Early results reveal that while duration-based effects are similar across groups, there may be differences in how musically trained and untrained listeners segment based on growl.

Induction and Prevention of Speech Learning

Beverly A. Wright, Department of Communication Sciences and Disorders, Northwestern University

Perceptual skills improve with practice. My coworkers and I have been investigating the factors that induce and those that prevent durable learning on speech-perception tasks (acquisition of a novel phonetic category and adaptation to foreign-accented speech). For example, we have evidence that: (1) for speech learning to survive across days requires a sufficient amount of training per day within a restricted time period, and that additional daily training can be superfluous; (2) a combination of practice with relevant speech stimuli and additional stimulus exposures without practice can enhance speech learning; (3) practice on more than one speech task within a single session can either enhance or disrupt learning depending on how the training trials are distributed; and (4) speech training regimens that yield learning reliably in young adults can be much less effective in adolescents. We have previously reported each of these learning patterns for fine-grained auditory discrimination tasks, suggesting that common principles are at play across speech and non-speech domains. Knowledge of these issues will lead to more effective training strategies to enhance speech learning in typical and clinical populations. Thanks, Randy, for your mentorship and wonderful sense of humor.

Nonlinear Auditory Models Yield New Insights into Representations of Vowels

Laurel H. Carney, Departments of Biomedical Engineering and Neuroscience, and Joyce M. McDonough, Department of Linguistics, University of Rochester

The potential benefit of understanding neural representations of speech sounds has long been recognized (e.g. Diehl, Lindblom, & Creeger, 2003). The representations of sound in waveforms or spectrograms are not an accurate model of the speech signal in the auditory pathway (e.g. Liljencrants and Lindblom, 1972; Carlson and Granström, 1978; Lindblom, 1986; Diehl et al., 2003). Many auditory models used to explore speech responses have been based on linear filterbanks, and do not incorporate key nonlinear response properties of the inner ear. The simple mapping of spectral properties into auditory responses by the responses of linear models is not realistic and, in particular, is not accurate at the sound levels of conversational speech. Detailed computational neural models for auditory-nerve fibers are available, with realistic nonlinear properties associated with the basilar membrane, inner hair cells, and auditory-nerve synapse (e.g. Zilany et al., 2014). These models reveal how the harmonic structure of vowels sets up patterns of strong F0-related amplitude fluctuations in auditory-nerve responses. The amplitudes of these fluctuations in auditory-nerve responses vary across frequency channels: the fluctuations are smallest near formant peaks, and largest at frequencies between formants (Carney, Li, McDonough, 2015). These F0-related fluctuations can strongly excite or suppress neurons in the auditory midbrain, the first level of the auditory pathway where tuning for low-frequency modulations in sounds occurs. Recent work has shown that the formant-related variations in amplitude fluctuations provide a representation of the vowel spectrum in the discharge rates of midbrain neurons that is robust across a wide range of sound levels and in the presence of background noise (Carney et al., 2015). These studies were based on neural tuning for amplitude-modulation frequency, a well-known property of midbrain neurons. More recently, strong selectivity of midbrain neurons to fast frequency transitions ("chirps") has been observed (Carney, Oetjen, Klump, 2017 ASA abstract). Chirps occur in harmonic sounds, such as vowels, due to the phase properties of vocal resonances, which affect the relative timing of the harmonics. The selectivity of midbrain neurons for chirps in harmonic sounds is consistent with the strong impact of phase manipulations on perception of vowel sounds (Carlson, Granström & Klatt, 1979). Realistic neural models for auditory-nerve and midbrain responses provide new insight into critical properties of vowel contrasts and their representation at these levels of the auditory pathway.

Who Knows Where the Time Went? Temporal Integration of Acoustic Input During Spoken Word Recognition is Not Always Linear

Bob McMurray, Kayleeen E. Schreiber, Jamie Klein, and Marcus E. Galle, Department of Psychological and Brain Sciences, University of Iowa

A classic issue in speech perception is how listeners integrate the multiple acoustic cues that signal phonological contrasts like voicing and place of articulation. Recently, a number of studies have pointed out the need to consider time. For most (if not all) phonetic contrasts, the relevant phonetic cues are not available at the same moment. For example, voicing is signaled by the voice onset time (VOT) at word onset, but also by the length of the following vowel, which may not be computable until 100 milliseconds later (or more). Consequently, integration requires the use of some form of memory or memory-like processes. In this regard, several studies report evidence that listeners make immediate partial commitments to multiple phonemic or lexical alternatives as soon as the first cue arrives. These commitments are then updated when later cues are heard (e.g., McMurray, Clayards, Aslin, Tanenhaus, 2008; Reinisch &Sjerps, 2013). More recently, we examined this in the context of sibilant fricatives (Galle, Klein-Packard, Schreiber & McMurray, submitted) which 1) are highly sensitive to contextual factors like coarticulation from the vowel; and 2) span a substantially different frequency band than other speech cues. We used the visual world paradigm to ask when the spectrum of the fricative is integrated with the formant transitions in the vocoid as well as with vowel identity (coarticulation). Across four experiments, we (quite surprisingly) found that listeners wait until the onset of the vocoid to use any of the information, implying some form of memory buffer. This could not be attributed to an idiosyncratic property of the stimuli or task, and even appeared in running speech. We argued that this pattern of integration (which deviates from what has been seen with all other speech cues) suggests that fricatives may be processed differently because their high frequency aperiodic nature cannot be as easily integrated (at an auditory level) with the rest of the speech. We then present new data suggesting a more complex form of integration. Here, we examined when listeners use coarticulation in the fricative to predict an upcoming vowel in words like seat and soup. A visual world study on this showed, that listeners could anticipate the vowel before they could identify the fricative. This suggests that it is not the aperiodic nature of the fricative that is driving the effect, but rather a need to identify the vowel, as a potentially organizing unit of the syllable. This is discussed within the broader context of studies of a range of phenomena which suggest a rather non-linear relationship between time in the input and the dynamics of cognitive processing.

Intelligibility Variation: Talkers, Listeners, and Signals

Rajka Smiljanic, Department of Linguistics, University of Texas, Austin

Understanding speech is implicit and automatic in favorable listening conditions (Rönnberg, 2003; Rönnberg et al., 2013). However, daily communication occurs in noise, in a foreign language, or with hearing loss. Under these circumstances, speech processing becomes more demanding and effortful. Processing an acoustically degraded or ambiguous signal requires listeners to engage more cognitive resources, leaving fewer of these resources for subsequent processing such as comprehension, storing linguistic information in memory, and recall (Hygge, Kjellberg, & Nöstl, 2015; Koeritzer, Rogers, Van Engen, & Peelle, 2018; Rabbitt, 1990; Rönnberg et al., 2013; Tun, McCoy, & Wingfield, 2009). In this talk, I discuss a series of experiments that explore the extent to which signal clarity, sentence context, and second language processing affect word recognition and sentence recognition memory. The results contribute evidence that listener-oriented speaking style modifications and semantic predictability enhance intelligibility and provide cognitive release to native and non-native listeners. Examining tasks that increase cognitive effort during listening provides important insights into how these cognitive demands are affected by talker-, listener-, and signal-related characteristics.