4pSC13. Auditory perception of sentence-level focus in Mandarin: A comparison of Korean learners and native speakers. Sun Hee Lee (Chinese, Cyber Hankuk Univ. of Foreign Studies, WORLDCUPRO 338, 303-1701, Seoul 03911, Korea, lishanxi@cufs.ac.kr)

Understanding spoken language requires more than recognizing individual sounds; prosodic features such as sentence-level focus play a key role in guiding listener attention and interpreting meaning. This study investigates how listeners' native language backgrounds influence their auditory perception of prosodic focus in Mandarin Chinese. We used the sentence 猫咪吃西瓜 ("The cat eats watermelon") and created four versions: ST1 with manipulated prosodic cues at the subject, ST2 at the verb, ST3 at the object, and ST4 as the unmodified baseline. Each focused version was manipulated using duration, pitch (F0), and intensity cues. Twelve native Mandarin speakers and twelve Korean learners of Mandarin participated in a perception task using PsychoPy, selecting the perceived focus position in each stimulus. Results revealed that both groups showed lower accuracy when only one acoustic cue was present, particularly in ST1. Accuracy improved when multiple cues were combined. Interestingly, native Mandarin listeners showed difficulty in detecting object focus (ST3), possibly due to a syntactic strategy in Mandarin that topicalizes objects rather than emphasizing them prosodically. These findings suggest that both language background and syntactic expectations shape focus perception. The study contributes to understanding how second-language learners and native listeners differ in processing prosodic cues during sentence comprehension.

4pSC14. Cross-linguistic effects of Korean experience on Chinese listeners' perception of English lexical stress. Eunkyung Sung (English, Cyber Hankuk Univ. of Foreign Studies, Imunro 107, Seoul, Korea, eks@cufs.ac.kr) and Sun Hee Lee (Chinese, Cyber Hankuk Univ. of Foreign Studies, Seoul, Korea)

This study explores how cross-language experience with Korean influences Chinese listeners' perception of English lexical stress, focusing on three prosodic cues: duration, fundamental frequency (F0), and intensity. Three groups participated: native Korean speakers (KL), Chinese speakers in Korea with Korean experience (CK), and Chinese speakers in China with no Korean experience (CC). Participants listened to English nonce words and identified the location of stress. All groups were relatively insensitive to duration, with the KL group showing the least sensitivity. In contrast, F0 and intensity cues strongly influenced perception, often leading to iambic interpretations. Notably, the CK group showed reduced sensitivity to F0 compared to both KL and CC groups. This suggests that experience with Korean, which encodes prominence differently than English, may shift cue weighting strategies in cross-language prosodic perception. Statistical analysis confirmed significant effects of F0 and intensity, along with interactions between listener group and cue type. These findings highlight how third-language experience can shape prosodic processing in a second language, contributing to models of cross-language speech perception.

4pSC15. How does attentional control modulate the younger and older adults' dichotic listening of Cantonese tones? Yuqi Wang (Hong Kong Univ. of Sci. and Technol., Rm. 3001, Academic Bldg., HKUST, Hong Kong, ywangqi@connect.ust.hk), Xiaohu Yang (Tongji Univ., Shanghai, China), and Zhen Qin (Hong Kong Univ. of Sci. and Technol., Hong Kong, Hong Kong)

Attentional control is found to modulate speech (i.e., consonants) processing in a forced-attention dichotic listening (FADL) task. A right-ear advantage (REA) in the non-forced condition (NF, baseline: hearing two stimuli simultaneously and identifying the clearer one) is often enhanced when participants are instructed to report the right-ear stimulus (forced-right, FR) or switched to a left-ear advantage (LEA) when instructed to report the left-ear stimulus (forced-left, FL). Different from consonants that are cued by voice onset time, lexical tones are often cued by pitch and found to show an LEA baseline. This study investigates whether younger adults show an effect of attentional control on tone processing and whether their older counterparts exhibit a less pronounced effect due to an age-related decline. We first tested 60 Cantonese-dominant younger participants (18–25 years) with the Cantonese tone FADL task under three conditions (NF, FL, FR) in Experiment 1. They increased the LEA baseline in the FL and showed the opposite REA pattern in the FR, suggesting an effect of attentional control. Experiment 2 tested 64 older participants (59–72 years), who showed an LEA baseline, a similar REA in the FR, but a smaller LEA increase in the FL than younger adults, suggesting a potential decline in attentional control. This decline appeared in one but not both forced conditions, implying tones—with more salient, longer-duration cues than consonants—may partially preserve auditory attentional control in aging.

4pSC16. Apparent talker variability and speaking style similarity facilitate comprehension of novel L2-accented talkers: Evidence for numerosity accounts of cross-talker generalization. Nicholas Aoki (Lingust., Univ. of California, Davis, 2100 5th Str. (Apt 131), Davis, CA 95618, nbaoki@ucdavis.edu) and Georgia Zellou (Lingust., Univ. of California, Davis, Davis, CA)

Talker variability can facilitate cross-talker generalization (e.g., comprehension of novel L2-accented talkers can increase after multi-talker exposure compared to single-talker exposure). However, current work contains a confound—do multi-talker exposure benefits stem from *numerosity* (greater number of talkers) or *heterogeneity* (greater phonological variability)? The current study assigned L1-English participants ($n = 815$) to one of five exposure conditions: (i) no exposure [control]; (ii) single-talker, casual speech; (iii) single-talker, hard-of-hearing-directed speech; (iv) multiple apparent talkers, casual speech; (v) multiple apparent talkers, hard-of-hearing-directed speech. There were two test conditions: (i) single-talker, casual speech; (ii) single-talker, hard-of-hearing-directed speech. Listeners transcribed Mandarin-accented English sentences in noise across both phases (a novel talker was always presented at test). Critically, participants only heard a single exposure talker in reality. *Apparent* talker variability was created through the "Change Gender" Praat function, which we claim to boost the number of perceived talkers while mitigating changes in phonological variation (i.e., an increase in numerosity with minimal adjustment in heterogeneity). We find that multiple apparent talker exposure can enhance generalization (greater test phase accuracy versus control), but only given acoustic similarity in speaking style across exposure and test.

4pSC17. Perceptual comparison of emotional Korean speech: Human versus AI-generated voices. Na-Young Ryu (Penn State Univ., University Park, PA) and Suyeon Yun (Chungnam National Univ., 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Korea, suyeon.yun@cnu.ac.kr)

This study examined how native Korean listeners perceive emotional speech produced by professional human voice actors versus speech generated by Zyphra's Zonos-v0.1, an advanced AI TTS system. 87 Korean participants listened to 96 audio stimuli, derived from 12 Korean sentences designed to evoke six core emotions and neutral expressions, each rendered by both human actors and AI voices cloned from those actors. Participants assessed each sample for voice identity (human or AI), naturalness, pronunciation, intonation, perceived emotion, and emotional appropriateness. Results revealed that while listeners accurately identified AI voices as artificial (78.81%), they often misclassified human voices as AI (53.90%), especially for neutral speech, indicating a narrowing perceptual gap. Voices perceived as human received higher ratings for naturalness, pronunciation, and intonation, regardless of their true origin. AI voices were rated lower, mainly due to less convincing emotional expression and intonation, though pronunciation was rarely an issue. Human voices were also much more likely to have their intended emotion correctly identified (65.54% vs. 31.66% for AI), and their emotional appropriateness was rated higher. In conclusion, while AI voices have advanced to be confused with human speech, limitations persist in achieving emotional authenticity and contextual appropriateness.

4pSC18. Neural processing of phonemes and allophonic variants across English accents. Yasuaki Shinohara (Faculty of Commerce, Waseda Univ., 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan, y.shinohara@waseda.jp), Ebony Goldman, Kevin Guzzo, and Valerie Shafer (The Graduate Ctr., City Univ. of New York, New York, NY)

This study examined how the brains of American and British English speakers responded to their respective allophonic variants of intervocalic /t/.