Title	Perception of Cantonese Parkinsonian speech
Author(s)	Wong, PCM; Diehl, RL; Ho, SL; Li, LSW; Tsang, KL
Citation	The 141st Annual Meeting of the Acoustical Society of America, Chicago, IL., 4-8 June 2001. In Acoustical Society of America Journal, 2001, v. 109 n. 5 pt. 2, p. 2314
Issued Date	2001
URL	http://hdl.handle.net/10722/54155
Rights	Creative Commons: Attribution 3.0 Hong Kong License

information. The present work investigates how temporal changes in three-dimensional distribution of early reflections influence speech intelligibility in rooms. A new measurement method, using a five microphone array and an omnidirectional source setup, is employed, and a series of post-processing procedures are involved, for getting different early reflections in their spatial distributions. The changes were made for the impulse responses obtained through a five microphone array in the arrival times of early reflections from all, and the horizontal and vertical directions, respectively. Anechoic samples of the Korean language were convolved binaurally with the reproduced impulses by applying a head-related transfer function. A series of speech intelligibility tests, conducted for 22 university students, found that the percentage of correct responses significantly deteriorated by increasing delay times of early reflections from the vertical direction. The result suggests that vertical components of early reflections play a significant role in improving speech intelligibility. [Work supported by Korean Research Foundation Grant KRF-1999-1-310-004-3.]

1pSC17. Consonants and vowels discriminated differently even when acoustically matched.

A. Min Kang (Haskins Labs., 270 Crown St., New Haven, CT 06511 and Yale Univ., New Haven, CT 06520, min.kang@yale.edu)

Vowels are reportedly discriminated differently from consonants, but there have typically been large between-class acoustic differences. Discrimination still differed when acoustic differences were reduced by removing the mostly vocalic center portion of CVCs [silent center (SC)] [A. M. Kang and D. H. Whalen, J. Acoust. Soc. Am. **107**, 2855–2856 (2000)]. The present study compared consonant and vowel identification and discrimination of synthetic CVCs varying in equal-sized $F2$ steps along /b-d/ and /ɛ-ʌ/ continua (full syllables), and in truncated syllables corresponding to the initial 60 ms of the previously examined SC syllables. To lower listener uncertainty, only consonant, or only vowel, information was varied within a test block. Consonant discrimination for full syllables was much higher than in the earlier SC experiment; it was slightly higher for the truncated stimuli than for the full. Vowel discrimination was much higher than consonant, near ceiling for both full and truncated stimuli. Thus, even when acoustic steps are equalized and the speech presented (in the truncated stimuli) is limited to the syllable portion that contains most of the constant information, vowels remain better discriminated than consonants. This indicates a true difference processing of the two phonetic classes, even when the acoustics are well matched. [Work supported by NIH.]

1pSC18. Perception of Cantonese Parkinsonian speech.

Patrick C. M. Wong, Randy L. Diehl (Dept. of Psych., Univ. of Texas, Austin, TX 78712), Shu Leong Ho, Leonard S. W. Li (Univ. of Hong Kong, Hong Kong, PROC), and Kin Lun Tsang (Queen Mary Hospital, Hong Kong, PROC)

The current study is a continuation of our previous case study investigating the effect of reduced pitch range in Parkinsonian speech on a tone language [P. C. M. Wong and R. L. Diehl, J. Acoust. Soc. Am. **105**, 1246(A) (1999)]. In the first experiment, listeners were asked to identify the last word of semantically neutral sentences produced by Cantonese-speaking Parkinson's disease (PD) patients, normal speakers, and a resynthesized version of PD speech with expanded pitch range. Identification of normal and PD speech did not differ, perhaps due to the insignificant difference in pitch range between the two types of speech. However, listeners were better at identifying the resynthesized PD speech which contained a larger pitch range than the original PD speech. This latter result supports the theory of context-target pitch distance proposed by Wong and Diehl which states that lexical tone perception relies on a sufficiently large pitch distance between the context and target of an utterance [J. Acoust. Soc. Am. **104**, 1834(A) (1998)]. In the second experiment, subjects were asked to identify the intended intonation (angry, happy, neutral, and question) of sentences produced by normal and PD speakers. Performance was better for normal speech. [Work supported by NIDCD.]

1pSC19. Do listeners to speech perceive gestures? Evidence from choice and simple response time tasks.

Carol A. Fowler, Julie M. Brown, Laura Sabadini-Grant (Haskins Labs. and Dept. of Psych., Univ. of Connecticut, 406 Babbidge Rd., Unit 1020, Storrs, CT 06269, fowler@tom.haskins.yale.edu), and Jeffrey Weihing (Haskins Labs., New Haven, CT 06511-6695)

According to the motor and direct realist theories, listeners perceive speech gestures. The following experiments test this claim. Experiments 1 and 2 replicate the findings of Porter and Castellanos [J. Acoust. Soc. Am. **67**, 1349–1356 (1980)]. Participants shadowed vowel-consonant-vowels (VCVs) produced by a model. Responses were timed. The difference between response times (RT) in simple and choice speech shadowing tasks (26 ms) is shorter than the canonical choice/simple RT difference [100–150 ms, Luce (Oxford, New York, 1986)]. This is interpreted as supporting Porter and Castellanos, in that when the task is to shadow speech, the element of choice is considerably reduced as the listener receives instructions for her response from the speech sounds she perceives. In experiment 3, the timing of gestures of the models' speech was manipulated by extending the voice onset time (VOT) of the models' production of voiceless stops in half of the speakers VCVs. VOTs of participants shadowed responses were measured. Our findings suggest that listeners' productions of phonemes can be influenced by their perception of the timing of the models' gestures in speech shadowing tasks. This provides additional support for the interpretation that participants' shadowing responses are guided by their preception of the models' gestures.

1pSC20. Bandpass filtered faces and audiovisual speech perception.

Kevin Munhall (Dept. of Psych. and Otolarngol., Queen's Univ., Kingston, Canada), Christian Kroos, and Eric Vatikiotis-Bateson (ATR Intl.—Information Sci. Div., Kyoto, Japan)

The visual system processes images in terms of spatial frequency-tuned channels. However, it is not clear how complex object and motion processing are influenced by this early visual processing. In two studies this question was explored in audiovisual speech perception. Subjects were presented with spatial frequency filtered images of the moving face during a speech in noise task. A wavelet procedure was used to create five bandpass filtered stimulus sets. The CID Everyday sentences were presented with a multivoice babble noise signal and key word identification accuracy was scored. Performance varied across the filter bands with peak accuracy being observed for the images containing spatial frequencies spanning 7–14 cycles/face. Accuracy for higher and lower spatial frequency bands was found to be lower. When viewing distance was manipulated no change in the overall shape or peak in the key word accuracy function was observed. However, at the longest viewing distance the performance in the highest spatial frequency band decreased markedly. The results will be discussed in terms of visual information processing constraints on audiovisual integration.

1pSC21. Neighborhood effects in Japanese word recognition.

Kiyoko Yoneyama and Keith Johnson (Dept. of Linguist., Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, yoneyama@ling.ohio-state.edu)

This paper reports on the results of a naming experiment that investigated lexical neighborhood effects in Japanese word recognition. A naming experiment was conducted with 28 Japanese adult listeners. Each participant responded to 700 words that had varying neighborhood density (in terms of Greenberg–Jenkins' phoneme substitution, deletion, and insertion rules). The lexicon used for this calculation consisted of only nouns from the NTT Japanese psycholinguistic database [Amano and Kondo (1999)]. A preliminary regression analysis showed that such neighborhood density was negatively correlated with naming reacting time. The words with higher neighborhood density were responded to faster than those with lower neighborhood density. We plan to report further analyses that (1)