



Title	An acoustic investigation of the Cantonese vowels in the speech of the adult and child speakers
Author(s)	Lee, WS
Citation	149th Meeting of the Acoustical Society of America, Vancouver, Canada, 16-20 May 2005, v. 117 n. 4 Pt 2, p. 2428-2429
Issued Date	2005
URL	http://hdl.handle.net/10722/54213
Rights	Creative Commons: Attribution 3.0 Hong Kong License

2aSC20. Method for automatic measurement of second language speaking proficiency. Jared Bernstein and Jennifer Balogh (Orinate Corp., Menlo Park, CA 94025)

Spoken language proficiency is intuitively related to effective and efficient communication in spoken interactions. However, it is difficult to derive a reliable estimate of spoken language proficiency by situated elicitation and evaluation of a person's communicative behavior. This paper describes the task structure and scoring logic of a group of fully automatic spoken language proficiency tests (for English, Spanish and Dutch) that are delivered via telephone or Internet. Test items are presented in spoken form and require a spoken response. Each test is automatically-scored and primarily based on short, decontextualized tasks that elicit integrated listening and speaking performances. The tests present several types of tasks to candidates, including sentence repetition, question answering, sentence construction, and story retelling. The spoken responses are scored according to the lexical content of the response and a set of acoustic base measures on segments, words and phrases, which are scaled with IRT methods or parametrically combined to optimize fit to human listener judgments. Most responses are isolated spoken phrases and sentences that are scored according to their linguistic content, their latency, and their fluency and pronunciation. The item development procedures and item norming are described.

2aSC21. Effects of linguistic experience on early levels of perceptual tone processing. Tsan Huang (Dept. of Linguist., SUNY Buffalo, 609 Baldy Hall, Buffalo, NY 14260) and Keith Johnson (UC Berkeley, Berkeley, CA 94720-2650)

This study investigated the phenomenon of language-specificity in Mandarin Chinese tone perception. The main question was whether linguistic experience affects the earliest levels of perceptual processing of tones. Chinese and American English listeners participated in four perception experiments, which involved short inter-stimulus intervals (300 ms or 100 ms) and an AX discrimination or AX degree-of-difference rating task. Three experiments used natural speech monosyllabic tone stimuli and one experiment used time-varying sinusoidal simulations of Mandarin tones. AE listeners showed psychoacoustic listening in all experiments, paying much attention to onset and offset pitch. Chinese listeners showed language-specific patterns in all experiments to various degrees, where tonal neutralization rules reduced perceptual distance between two otherwise contrastive tones for Chinese listeners. Since these experiments employed procedures hypothesized to tap the auditory trace mode (Pisoni, *Percept. Psychophys.* **13**, 253–260 (1973)), language-specificity found in this study seems to support the proposal of an auditory cortical map [Guenther *et al.*, *J. Acoust. Soc. Am.* **23**, 213–221 (1999)]. But the model needs refining to account for different degrees of language-specificity, which are better handled by Johnsons (2004, *TLS03:26-41*) lexical distance model, although the latter model is too rigid in assuming that linguistic experience does not affect low-level perceptual tasks such as AX discrimination with short ISIs.

2aSC22. Perceptual and production variables in explicating interlanguage speech intelligibility benefit. Ameer P. Shah and Zoi Vavva (Dept. of Speech and Hearing, Cleveland State Univ., 2121 Euclid Ave., MC 431-B, Cleveland, OH 44115, a.shah101@csuohio.edu)

This study attempts to investigate the importance of the degree of similarity or difference in the language backgrounds of the speakers and listeners, as it interacts differentially in intelligibility judgment of foreign-accented speech (Bent and Bradlow, 2003). The present study attempts to clarify the distinction in the matched and mismatched listening conditions, in context of addressing the overarching question whether auditory exposure to a language alone, without corresponding proficiency in production of that language, can provide a listening advantage. Particularly, do listeners understand accented-English speech spoken by native individuals of the language to which they are exposed to, as compared to listeners without that exposure? Greek-accented English speakers (and native monolin-

gual English speakers) were judged for their speech intelligibility by four groups of listeners ($n=10$, each): native Greek speakers (matched), Greek-Americans (matched only through auditory exposure to Greek without any corresponding spoken proficiency), native monolingual American-English speakers (unmatched), and a mixed group (mismatched). Pilot data have shown that the intelligibility judgments by Greek-American listeners are intermediate to the native Greeks, and both the American-English and the mixed group. Further data-collection is underway, and will be presented as they bear important theoretical and clinical implications.

2aSC23. Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. Kyoung-Ho Kang and Susan G. Guion (Dept. of Linguist., 1290 Univ. of Oregon, Eugene, OR 97403, kkang@darkwing.uoregon.edu)

The stop systems of adult Korean (L1)-English (L2) bilinguals were studied through acoustic analysis of Korean and English stop productions. The interaction of the first and second language stop systems was investigated as a function of age of exposure to English. The main goals of the investigation were to examine the extent to which early and the late bilinguals produced a given stop category in a native-like way and the extent to which the two stop systems were held independent from each other in the phonological systems of the bilinguals. Two specific questions were asked: whether early bilinguals were more native-like in the production of English stops and thus were more likely to establish L2 phonetic categories than late bilinguals, and whether the Early bilinguals maintain a greater extent of independence between the Korean and English stop systems than the late bilinguals. For this purpose, productions of Korean and English stops were analyzed in terms of three acoustic-phonetic properties, VOT, $H1-H2$, and f_0 . The results suggested that fine-grained phonetic information is crucial to investigation of bilingual phonological systems and they were discussed in terms of the role of perceived phonetic distance between L1 and L2 phonetic categories in L2 speech learning.

2aSC24. Validation of an automatic measurement of Spanish speaking proficiency. Elizabeth Rosenfeld, Jared Bernstein, and Jennifer Balogh (Orinate Corp., Menlo Park, CA)

A 15-min computer-based test of spoken Spanish was designed to measure candidate proficiency in Spanish. The test presents seven tasks: reading, elicited imitation, word opposites, short-answer questions, sentence constructions, opinion questions, and story retellings. The tests were presented to 579 adult non-native Spanish learners and to 552 native Spanish speakers. Expert human judgments of the non-native responses showed that the spoken response material carried sufficient information for highly reliable judgments of proficiency. In the development and validation process, 57 000 responses were transcribed and 21 000 human judgments were analyzed. The paper describes the validation of the automatic scoring system with reference to concurrent oral proficiency interviews conducted by professional raters certified by the US Government or by ACTFL. The outcomes of the comparisons of the machine scored tests with interactive human interviews and with human ratings from recorded speech indicate that the test produces scores that have virtually the same information that is found in oral proficiency interviews. Almost all assessments correlate highly with the other assessments with coefficients in the range 0.86–0.96. The test correlation with the combined interview scores ($r=0.92$) is higher than the inter-rater reliability of the professional interviewers themselves.

2aSC25. An acoustic investigation of the Cantonese vowels in the speech of the adult and child speakers. Wai-Sum Lee (Dept. of Linguist., The Univ. of Hong Kong, Pokfulam Rd., Hong Kong, wsleeba@hku.hk)

The study analyzes the formant center frequencies for the seven Cantonese vowels [i, y, u, e, æ, ɔ, a] from 30 native speakers of Cantonese, 10 male and 10 female adults and 5 male and 5 female 9–10 year old children. Results show that the formant frequencies for the vowels are

largest for the female children, followed by the male children, female adults, and male adults in decreasing order. Despite the differences, the patterns of formant frequencies for any one vowel for the different groups are similar. The difference in F -values for any one vowel between the male and female children is smaller than the difference between the male and female adults. As for individual formant frequencies, the difference in $F1$ between the males and females of the same age group and between the adults and children of the same gender group is smaller for the high vowels [i, y, u] than the non-high vowels [ɛ, æ, ɔ, a]. The difference in $F2$ between the males and females of the same age group and between the adults and children of the same gender group is smaller for the high rounded vowels [y, u] than the other vowels. The paper will also present the ratios of speaker group-to-speaker group for individual formant frequencies.

2aSC26. Voice onset time (VOT) in Canadian French and English: Monolingual and bilingual adults. Andrea A. N. MacLeod and Carol Stoel-Gammon (Dept. of Speech and Hearing Sci., Univ. of Washington, 1417 N.E. 42nd St., Seattle, WA 98119)

This study focused on the contrasts produced by early bilingual speakers ($n=6$) across their two languages in comparison with monolingual speakers (Canadian English (CE), $n=5$; Canadian French (CF), $n=6$). VOT production was measured in monosyllabic CE and CF words that began with one of four stop consonants, /p, b, t, d/ followed by one of three vowels. A total of 14–18 words for each of the four stop consonants for each language was elicited with a total number 1700 acoustically analyzed productions. The participants were tested individually in quiet rooms using a single target language throughout the session. As expected, the monolingual speakers produced a two-way contrast (statistically significant: $p < 0.05$): for CE speakers, short-lag VOT versus long-lag VOT; for CF speakers, lead VOT versus short-lag VOT. Rather than producing a two-way contrast (e.g., lead VOT versus lag VOT) or a three-way contrast (e.g., lead VOT versus short-lag VOT versus long-lag VOT), the bilingual speakers produced a four-way contrast (statistically significant: $p < 0.05$): long lead VOT (CF /b, d/), short lead VOT (CE /b, d/), short-lag VOT (CF /p, t/) and long-lag VOT (CE /p, t/). These results suggest that bilinguals are maintaining phonetic contrasts both within and across their two languages.

2aSC27. The effect of speaking rate on perception of syllables in second-language speech. Keiichi Tajima (Dept. of Psych., Hosei Univ., Tokyo 102-8160, Japan; ATR Human Information Sci. Labs., Kyoto 619-0288, Japan) and Reiko Akahane-Yamada (ATR Human Information Sci. Labs., Kyoto 619-0288, Japan)

Past studies on second-language (L2) speech perception have suggested that L2 learners have difficulty exploiting contextual information when perceiving L2 utterances, and that they exhibit greater difficulty than native listeners when faced with variability in temporal context. The present study investigated the extent to which native Japanese listeners, who are known to have difficulties perceiving English syllables, are influenced by changes in speaking rate when asked to count syllables in spoken English words. The stimuli consisted of a set of English words and non-words varying in syllable structure spoken at three rates by a native English speaker. The stimuli produced at the three rates were presented to native Japanese listeners in a random order. Results indicated that listeners' identification accuracy did not vary as a function of speaking rate, although it decreased significantly as the syllable structure of the stimuli became more complex. Moreover, even though speaking rate varied from trial to trial, Japanese listeners' performance did not decline compared to a condition in which the speaking rate was fixed. Theoretical and practical implications of these findings will be discussed. [Work supported by JSPS and NICT.]

2aSC28. Learning to talk: A non-imitative account of the replication of phonetics by child learners. Piers Messum (Dept. of Phonet. and Linguist., UCL, Gower St., London WC1E 6BT, UK)

How is it that an English-speaking 5-year-old comes to: pronounce the vowel of seat to be longer than that of sit, but shorter than that of seed; say a multi-word phrase with stress-timed rhythm; aspirate the /p/s of pin, polite, and spin to different degrees? These are systematic features of English, and most people believe that a child replicates them by imitation. If so, he is paying attention to phonetic detail in adult speech that is not very significant linguistically, and then making the effort to reproduce it. With all the other communicative challenges he faces, how plausible is this? An alternative, non-imitative account of the replication of these features relies on two mechanisms: (1) emulation, and (2) the conditioning of articulatory activity by the developing characteristics of speech breathing. The phenomena above then become no more than expressions of how a child finds ways to warp his phonetic output in order to reconcile conflicting production demands. The criteria he uses to do this make the challenges both of learning to talk and then of managing the interaction of complex phonetic patterns considerably more straightforward than has been imagined.

2aSC29. Acoustics of contrastive prosody in children. Rupal Patel, Jordan Piel (Dept. of Speech Lang. Pathol. & Audiol., Northeastern Univ., 360 Huntington Ave., Boston, MA 02115, r.patel@neu.edu), and Maria Grigos (New York Univ., New York, NY 10003)

Empirical data on the acoustics of prosodic control in children is limited, particularly for linguistically contrastive tasks. Twelve children aged 4, 7, and 11 years were asked to produce two utterances "Show Bob a bot" (voiced consonants) and "Show Pop a pot" (voiceless consonants) 10 times each with emphasis placed on the second word (Bob/Pop) and 10 times with emphasis placed on the last word (bot/pot). A total of 40 utterances were analyzed per child. The following acoustic measures were obtained for each word within each utterance: average fundamental frequency (f_0), peak f_0 , average intensity, peak intensity, and duration. Preliminary results suggest that 4 year olds are unable to modulate prosodic cues to signal the linguistic contrast. The 7 year olds, however, not only signaled the appropriate stress location, but did so with the most contrastive differences in f_0 , intensity, and duration, of all age groups. Prosodic differences between stressed and unstressed words were more pronounced for the utterance with voiced consonants. These findings suggest that the acoustics of linguistic prosody begin to differentiate between age 4 and 7 and may be highly influenced by changes in physiological control and flexibility that may also affect segmental features.

2aSC30. Infant-directed speech: Final syllable lengthening and rate of speech. Robyn Church, Barbara Bernhardt (School of Audiol. and Speech Sci., Univ. of British Columbia, Vancouver, BC, Canada V6T 1Z1), Rushen Shi (Univ. of Quebec at Montreal), and Kathleen Pichora-Fuller (Univ. of Toronto)

Speech rate has been reported to be slower in infant-directed speech (IDS) than in adult-directed speech (ADS). Studies have also found phrase-final lengthening to be more exaggerated in IDS compared with ADS. In our study we asked whether the observed overall slower rate of IDS is due to exaggerated utterance-final syllable lengthening. Two mothers of preverbal English-learning infants each participated in two recording sessions, one with her child, and another with an adult friend. The results showed an overall slower rate in IDS compared to ADS. However, when utterance-final syllables were excluded from the calculation, the speech rate in IDS and ADS did not differ significantly. The duration of utterance-final syllables differed significantly for IDS versus ADS. Thus, the overall slower rate of IDS was due to the extra-long final syllable occurring in relatively short utterances. The comparable pre-final speech rate for IDS and ADS further accentuates the final syllable lengthening in IDS. As utterances in IDS are typically phrases or clauses, the particularly