

**EFFECTS OF ENERGETIC AND INFORMATIONAL MASKING
ON SPEECH SEGMENTATION BY NATIVE AND NON-NATIVE SPEAKERS**

Sven L. Mattys, Lucy M. Carroll, Carrie K. W. Li, & Sonia L. Y. Chan

University of Bristol

Corresponding author:
Sven L. Mattys
University of Bristol
Department of Experimental Psychology
12A Priory Road
Bristol, BS8 1TU, UK
Sven.Mattys@bris.ac.uk

Abstract

In this study, we asked whether native and non-native speakers of English use a similar balance of lexical knowledge and acoustic cues, e.g., juncture-specific allophones, to segment spoken English, and whether the two groups are equally affected by energetic masking (a competing talker) and by cognitive load (a simultaneous visual search task). In intact speech, as well as in both adverse conditions, non-native speakers gave relatively less weight to lexical plausibility than to acoustic cues. Under energetic masking, overall segmentation accuracy decreased, but this decrease was of comparable magnitude in native and non-natives speakers. Under cognitive load, native speakers relied relatively more on lexical plausibility than on acoustic cues. This lexical drift was not observed in the non-native group. These results indicate that non-native speakers pay less attention to lexical information—and relatively more attention to acoustic detail—than previously thought. They also suggest that the penetrability of the speech system by cognitive factors depends on listener's proficiency with the language, and especially their level of lexical-semantic knowledge.

Key words: spoken-word recognition; speech segmentation; bilingualism; processing load; cognitive load; energetic masking; informational masking

The effect of adverse conditions on speech recognition is often investigated using intelligibility measures such as sentence transcription and keyword recognition, with the magnitude of the decrease in accuracy interpreted as reflecting the severity of the adverse condition. A complementary approach to the issue was recently proposed by Mattys, Brooks, and Cooke (2009), who assessed the effect of adverse conditions on listeners' relative reliance on lexical-semantic information and acoustic-phonetic cues, independent of intelligibility. A relative measure of knowledge-driven and signal-driven processes is particularly relevant to the study of speech segmentation, which is often seen as the result of a balance between the two (e.g., Davis, Marslen-Wilson, & Gaskell, 2002; Mattys, White, & Melhorn, 2005; Norris, McQueen, & Cutler, 1995). An understanding of how adverse listening conditions influence such a balance is central to current models of speech recognition not only because it can increase the external validity of these models, but also because it can provide an insight into the link between language processing and cognition.

Drawing from key notions in psychophysics and hearing science, Mattys et al. (2009) made a distinction between adverse conditions leading to energetic masking and those leading to informational masking (see, e.g., Brungart, 2001, for a review). Energetic masking occurs when portions of a signal are physically degraded by a distractor (e.g., background noise) in the same spectro-temporal regions. The target signal can then only be inferred through "glimpses" between the masked regions (Barker & Cooke, 2007; Cooke, 2006). Informational masking corresponds to the higher-level consequences of adverse conditions once energetic masking, if present, has been taken into account. Following Cooke, Garcia Lecumberri, and Barker's (2008) taxonomy, Mattys et al. divided informational masking into three categories: (1) Competing attention of the masker, i.e., the cost of the effort involved in ignoring the mask—whatever it is—through stream segregation or selective attention, e.g., speech in a background of babble noise; (2) Interference from a known language, that is, the detrimental effect of a mask when the mask itself is intelligible and meaningful. This kind of masking is thought to be due to lexical-semantic interference between mask and target, e.g., a intelligible competing talker; (3) Cognitive load, i.e., the depletion of processing resources incurred when listeners are required to divide their attention between the main task and the mask (e.g., attending to the competing talker's utterance) or between the main task and a secondary task (e.g., a visual search task).

To assess the effect of these various loads on speech segmentation, Mattys et al. (2009) devised a paradigm which provided a measure of listeners' relative reliance on lexical-semantic knowledge and acoustic cues on a single sliding scale. Listeners heard two-word phrases (e.g., "mild option"), whose acoustic realization ranged from being compatible with the lexically acceptable parse (e.g., "mild#option," with # representing the acoustically favored boundary) to being in conflict with it (e.g., "*mile#doption," with * denoting a lexically unacceptable segmentation solution). Listeners were asked to rate on an 11-point scale which of two words they heard (e.g., "mild" or "mile"). In this example, responding "mild" was taken as an indication that the listeners adhered predominantly to the lexical-semantic structure of the phrase, whereas responding "mile" (especially at the "*mile#doption" end of the range) was taken as an indication that the listeners relied predominantly on the acoustic cues.

Using this task, Mattys et al. (2009) reported four main findings. First, severe energetic masking (e.g., background multi-talker babble) led to relatively greater reliance on salient acoustic cues than on lexical-semantic plausibility, probably because lexical access and semantic integration were too compromised to be effective, hence causing listeners to fall back on any local acoustic cues that could be glimpsed from the signal. Second, the presence of a load, energetic or informational, always resulted in decreased segmentation accuracy, independent of a change in balance between acoustic and lexical-semantic reliance. Mattys et

al. attributed this effect to the cost involved in segregating the target from the distractor (Cooke et al.'s [2008] "competing attention of the masker"). Third, there was no evidence of "interference from a known language": An intelligible distractor, e.g., an English sentence, was no more distracting than its speech-shaped noise equivalent. Fourth, a cognitive load, such as that created by a simultaneous task, caused listeners to rely relatively more on lexical-semantic information than acoustic cues.

The above results were found for native speakers of English. However, there are reasons to believe that patterns might differ for non-native speakers. Indeed, even in optimal conditions, non-native speakers have been shown to assign different weights to segmentation cues compared to native speakers. For instance, Sanders and colleagues (Sanders & Neville, 2003; Sanders, Neville, & Woldorff, 2002) found that, while the use of lexical-semantic information for segmentation is comparable in native and non-native speakers of English, the use of syntactic information is attenuated in the non-native speakers. Likewise, the interpretation of word-juncture allophonic English variants (e.g., "keeps talking" vs. "keep stalking") is generally less accurate in non-native than native speakers (e.g., Altenberg, 2005; Ito & Strange, 2009).

Studies of sentence and phoneme intelligibility in noise have shown similar contrasts. In particular, a widespread observation is that the detrimental effect of background noise on sentence intelligibility is more pronounced for non-native than native speakers (e.g., Mayo, Florentine, & Buus, 1997; Nábělek & Donahue, 1984; Rosenhouse, Haik, & Kishon-Rabin, 2006). An "upstream" explanation suggests that the greater deficit in non-native speakers is due primarily to underspecified representations of non-native segments. In this view, it is the incomplete or erroneous encoding of segments that leads to a cascaded failure in lexical access and sentential integration, rather than inadequate lexical or syntactic knowledge per se (e.g., Bradlow & Alexander, 2007; Bradlow & Bent, 2002; Hazan & Simpson, 2000). An alternative explanation is that the non-native disadvantage is due just as much—if not more—to inadequate/incomplete lexical and syntactic knowledge as to poor segmental representations (Mayo et al., 1997; van Wijngaarden, Steeneken, & Houtgast, 2002). In this case, non-native speakers' unfamiliarity with low-frequency words, syntactic structures and idiomatic expressions would make compensation for sensory degradation more arduous. The contribution of lexical and post-lexical knowledge to non-native's low performance in noise is evidenced by the lack of interaction between native language and noise when the stimuli are devoid of lexical content or syntactic structure (Cutler, Garcia Lecumberri, & Cooke, 2008; Cutler, Weber, Smits, & Cooper, 2004).

Finally, laboratory research and research on speech communication between air-traffic controllers and pilots has shown that speech production and comprehension performance is broadly more affected by a cognitive load (e.g., a mental arithmetic task) in non-native than native speakers (Farris, Trofimovich, Segalowitz, & Gatbonton, 2008; Takano & Noda, 1993). Whether and how this pattern applies to the relative use of lexical-semantic information and acoustic cues for segmentation is unclear.

The above studies suggest that the patterns of segmentation strategies exhibited by native speakers in conditions of energetic and informational masking are unlikely to hold, at least in full, for non-native speakers. In this article, we attempted to answer four questions.

1. When the listening environment is devoid of any load, is the balance of weights between lexical-semantic and acoustic cues similar in native and non-native speakers? Three scenarios are possible: (a) Same balance of weights. Despite possible differences in how fully non-native speakers exploit the available segmentation cues, relative reliance would be comparable in both groups; (b) Relatively greater reliance on lexical-semantic structure for non-native speakers. This is expected if non-native speakers have poor representations

of word-juncture allophonic variants in the non-native language (Altenberg, 2005; Ito & Strange, 2009) and if higher-order knowledge is comparatively good (Sanders & Neville, 2003; Sanders et al., 2002); (c) Relatively greater reliance on acoustic cues for non-native speakers. This pattern is expected if poor lexical-semantic representations cause non-native speakers to solve ambiguous segmentation by falling back on lower-level sensory information (Mattys et al., 2005).

2. Are native and non-native speakers' segmentation strategies differentially affected by energetic masking? If the disproportionate effect of noise on speech intelligibility by non-native speakers (e.g., Nábělek & Donahue, 1984) is due to their inadequate segmental and allophonic knowledge, a compensatory drift in favor of lexical-semantic information could be expected, especially when the stimuli are lexically constrained by the design, as they will be in the present segmentation experiments. However, if energetic masking impairs lexical-semantic retrieval and integration, a compensatory drift in favor of acoustic cues—those than can be glimpsed—would be expected (Cutler et al., 2004, 2008; Mattys et al., 2005).
3. Does the informational content of a load affect the balance between lexical-semantic and acoustic reliance? Multi-talker babble in a known language has been shown to be more deleterious than multi-talker babble in an unknown language in both sentence intelligibility tasks (Van Engen & Bradlow, 2007) and phoneme identification tasks (Garcia Lecumberri & Cooke, 2006). However, Mattys et al. (2009) found no difference between an intelligible competing talker and its speech-shaped equivalent with their segmentation task. In the following experiments, we will test the linguistic interference hypothesis by using Mattys et al.'s procedure with native and non-native speakers and native and non-native competing talkers.
4. Does a cognitive load lead to relatively greater reliance on lexical-semantic information in both native speakers (as in Mattys et al., 2009) and non-native speakers? Mattys et al. found that native English speakers relied more heavily on lexical-semantic information than on acoustic cues when their attention was divided between the segmentation task and a concurrent task. They interpreted this lexical drift as evidence that reliance on low-level cues was more vulnerable to a depletion of central processing resources than reliance on higher-order information. There is no reason to believe that non-native speakers should respond differently unless they have initial limitations with processing lexical-semantic information.

The following experiments were based on the stimuli, design, and procedure of Mattys et al. (2009), which are summarized in the Method section of Experiment 1. The two populations of interest were native English speakers with no knowledge of Cantonese and native Cantonese speakers proficient in second-language English. We chose Cantonese as the non-native language because it offers a sharp contrast with English with respect to prosodic structure and phonotactic regularities, and because the two languages share few loan words. Cantonese was chosen over languages with similar characteristics because of the availability of Cantonese speakers in the local area.

Experiment 1

In this experiment, the English and Cantonese participants heard various acoustic renditions of two-word test phrases (e.g., "mild option") in three conditions: intact speech, against the background of a competing talker speaking English, and against the background of a competing talker speaking Cantonese.

Method

Participants.

Thirty-three native speakers of English and 33 native speakers of Cantonese with no self-reported hearing problems were recruited for this experiment. They received course credit or a small honorarium for their participation. The English speakers were recruited mostly from the undergraduate population at the University of Bristol. Most of them were between 18 and 25 years of age. The Cantonese individuals were proficient speakers of English as a second language, recruited from both the university undergraduate population and outside the university. Their age ranged from 20 to 23. All of them had lived in the UK for at least three years (average: 6 years; range: 3 to 10 years) and had studied English in school since they were six or under. Their English, though accented, was fluent. On a scale from 1 (very poor) to 5 (excellent), they rated their English pronunciation 3.6 on average (SD = .68) and their proficiency in English grammar 3.5 (SD = .63).

Materials.

Test phrases

A full description of the test phrases, which were produced by a female native speaker of British English, is available in Mattys et al. (2009). In brief, they consisted of three sets of two-word phrases varying in their acoustic realization. For example, "mild option" was produced so as to induce the perception of "mild" at the beginning, leading to the lexically acceptable parse "mild#option." These phrases were therefore called "lexical phrases." "Mild option" was also produced so as to induce the perception of "mile" at the beginning, leading to the lexically unacceptable parse "*mile#dooption." These phrases were called "acoustic phrases" because they contained acoustic cues strong enough to counteract the effect of lexical structure. In the third set, "mild option" was produced so as to induce a relatively ambiguous percept between "mild#option" and "*mile#dooption." These phrases were called "ambiguous phrases." For counterbalancing purposes, each phrase was matched with another phrase in which the segmentation of the long, rather than short target word yielded a non-lexical residue (e.g., "*mild#eep" vs "mile#deep"). Likewise, there was an equal number of phrases containing the test words phrase-initially (cf. "mild option" and "mile deep" examples) and phrase-finally (e.g., "sleep" and "leap" in "deep#sleep" vs. "*deeps#leap" and in "*clo#sleep" vs. "close#leap"). The lexical frequency of the two words within a phrase was matched (Celex; Baayen, Piepenbrock, & Gulikers, 1995) and the phonotactic properties of the juncture diphones were controlled. In all, there were 40 lexical phrases, 40 acoustic phrases, and 40 ambiguous phrases, for a total of 120 phrases (see Appendix A in Mattys et al.).

These stimuli were selected after piloting a large number of natural renditions produced in the laboratory. After each rendition, participants indicated which of two words they heard at the beginning (or the end) of the phrase, using an 11-point scale shown as eleven adjacent keys on a computer keyboard. A value of 10 was given to the word leading to a lexically acceptable segmentation, regardless of the length of the word or position of the word in the phrase. For instance, the "mild" answer was rated 10 in the "mild#option" – "*mile#dooption" phrases; "mile" was rated 10 in the "*mild#eep" – "mile#deep" phrases; "sleep" was rated 10 in the "deep#sleep" – "*deeps#leap" phrases; "leap" was rated 10 in the "*clo#sleep" – "close#leap" phrases. A value of 0 was given to the other word.

The average ratings for the lexical, ambiguous, and acoustic test phrases were, respectively: 7.13 (phrase range: 5.57-8.07), 5.04 (3.57-6.93), and 2.90 (2.07-4.21). Note that even the lexical and acoustic phrases were selected so as to contain a certain degree of ambiguity (i.e., we did not chose phrases with ratings near 10 or near 0) to allow the effects of energetic and informational masking to manifest themselves in those conditions as well. To verify that the lexical asymmetry in the phrases was indeed reflected in the ratings, as intended, ratings of the

phrases were also collected once the lexical asymmetry was removed. To so do, we spliced off some material at the end or beginning of the phrases. For example, "mild#option" became "*mild#opsh," "*mile#doption" became "*mile#dopsh," "deep#sleep" became "*eep#sleep," "*deeps#leap" became "*eeps#leap," etc. Truncation made all renditions lexically unacceptable, but it did not alter the acoustic juncture cues. The average ratings for the truncated lexical, ambiguous, and acoustic phrases were, respectively: 5.19 (1.15-7.40), 3.63 (0.95-7.05), and 2.07 (0.60-5.80). The rating drift in the acoustic direction confirms the impact of lexical structure on segmentation and the fact that our phrases did indeed contain various degrees of tension between lexical structure and acoustic cues, with a predominance of lexical structure in the lexical phrases, a predominance of acoustic cues in the acoustic phrases, and balanced tension between lexical structure and acoustic cues in the ambiguous phrases.

Acoustic analyses of the segments around the junctures in the three sets of phrases revealed quantitative and qualitative differences in line with established patterns in experimental phonetics, namely, word-initial consonant lengthening (e.g., Oller, 1973) and word-initial vowel glottalization (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996). More details can be found in Appendix B in Mattys et al. (2009).

Competing talkers

The materials for the English competing talker came from Experiment 2 in Mattys et al. (2009). They consisted of connected, meaningful utterances produced by a male native speaker of English. The utterances were extracts from short stories, with silence between words or phrases removed. For the purpose of this study, comparable utterances were recorded from a male native speaker of Cantonese, who read extracts from articles in Cantonese newspapers and magazines. Using male voices as competing talkers allowed us to make the instructions to participants easy to understand (see Procedure).

In both conditions, the competing utterance started 781 ms before the onset of the target phrase and ended 781 ms after its offset. The intensity of the competing utterances was scaled so that they caused comparable levels of energetic masking in both conditions. To estimate energetic masking, we measured the percentage of each test phrase that could be *glimpsed* through its paired competing utterance (Cooke, 2006), that is, the percentage of spectro-temporal regions where the test phrase had a local signal-to-noise ratio (SNR) in excess of the competing utterance. Because the density of gammatone filters along the frequency spectrum (50-8000 Hz) followed the inverse of a modified logarithmic function, low frequencies had relatively more weight than higher frequencies in the glimpsing calculation.

In Mattys et al., the energetic masking generated by the English competing utterances allowed 36 % of the spectro-temporal plane to be glimpsed. The SNR, measured over the overlap between the phrases and the competing talker, was -8 dB. In order to achieve a comparable glimpse percentage for the Cantonese competing utterances (37 %), the SNR had to be adjusted to -10.5 dB. We equated energetic masking based on glimpse percentage rather than SNR because, in contrast to SNR, which is not a very accurate predictor of intelligibility when different types of maskers are compared (Festen & Plomp, 1990), glimpse percentage is highly-correlated with intelligibility (Cooke, 2006).

Procedure.

For counterbalancing purposes, the 120 phrases were divided into three sets of 40. Twenty phrases of each set contained the target word phrase-initially and the other 20 phrase-finally. Each subset of 20 phrases contained 10 phrases in which the lexically acceptable parsing arose from the long word (e.g., "mild#option") and 10 in which it arose from the shorter word (e.g., "mile#deep"). Within each set of 10 phrases, 3 (or 4) were lexical, 3 (or 4) were ambiguous, and 3 (or 4) were acoustic. Which condition received 4 instead of 3 phrases varied across the four

sets of 10 phrases. The assignment of the three conditions (intact, English competitor, Cantonese competitor) to the three sets of 40 phrases was counterbalanced across triplets of participants in a Latin-square fashion. Thus, each participant was exposed to all three conditions, but never on the same phrases.

Each participant heard 120 phrases. Position of the target (initial, final) was blocked and counterbalanced across participants. Intact phrases, phrases in a competing English-talker background, and phrases in a competing Cantonese-talker background were mixed and randomized within each block. Each block was preceded by four practice trials. The phrases were played at 16 bit D/A, 32 kHz over SONY MDR-V700 headphones. The sound level of the intact phrases was kept at 65 dB SPL across all conditions.

The testing procedure was the same as in the pilot study. After hearing a phrase, participants reported which of two words (e.g., "mild" or "mile") they heard at the beginning of the phrase (or at the end, depending on the block). They were instructed to respond based on what they heard, not on what they thought the speaker should have said. The two words were shown on the left and right sides of a computer monitor, separated by dots. Participants gave their response using an 11-point scale shown as eleven adjacent keys on a computer keyboard. The leftmost keys represented adherence to the word on the left and the rightmost keys to the word on the right. The left/right position of the two words was counterbalanced across participants. Participants had up to 10 s to indicate their rating. They were also told that "Although the sequences will often be played in the clear, they will sometimes be played in the background of a distracting male voice speaking in English or Cantonese. Try to ignore the male voice as much as possible."

Results

The coding of the ratings was the same as in the pilot study, with a rating of 10 given to the word leading to a lexically acceptable segmentation and 0 to the word leading to a lexically unacceptable segmentation. Figure 1A shows the average ratings of the lexical, ambiguous, and acoustic phrases in intact speech, competing English talker, and competing Cantonese talker for the native English speakers. Figure 1B shows the same breakdown for the Cantonese speakers. Figures 1C and 1D are condensed displays of the same data, comparing English and Cantonese speakers directly. Specifically, Figure 1C shows the difference (spread) between the lexical and acoustic conditions. The spread is an estimate of the general cost of the load on segmentation judgements, with a small spread indicating poorer discrimination between segmentation alternatives. Figure 1D, in which the ratings of the three phrase types are averaged, shows the main effect of load on segmentation strategy, with an upward departure from baseline reflecting relatively greater reliance on lexical-semantic structure (which we refer to as a "lexical drift") and a downward departure from baseline reflecting relatively greater reliance on acoustic cues ("acoustic drift").

We first assessed the general cost of the competing talker on segmentation based on the spread data (Figure 1C). This was done by focusing on the interaction between, on the one hand, Phrase (restricted to the lexical and acoustic phrases) and, on the other hand, native language, i.e., L1 (English, Cantonese) and Load (intact, English competitor, Cantonese competitor), using a mixed-effect model (Baayen, Davidson, & Bates, 2008), with participants and items as random factors and Phrase, L1, and Load as fixed factors. A significant Phrase-by-L1 interaction, $F(1, 5254) = 7.68, p < .01$, indicated that the spread was smaller in the Cantonese group than in the English group. Thus, generally speaking, the non-native speakers were less discriminant than were the native speakers. This disadvantage was not affected by Load, as the three-way interaction was $F(2, 5254) < 1$. An interaction between Phrase and Load, $F(2, 5254) = 31.04, p < .001$, showed that the spread was smaller in the English-competitor and

Cantonese-competitor conditions than in the intact condition, $F(1, 3509) = 40.97, p < .001$, and $F(1, 3509) = 50.85, p < .001$, respectively. There was no difference between the two competitors, $F(1, 3509) < 1$. Thus, the presence of a competing talker was detrimental to listeners' overall segmentation performance, but the language of the competing talker made no difference to either the English or Cantonese listeners.

We then explored the effects of L1 and Load on the aggregated ratings (Figure 1D), averaging all three types of Phrase (lexical, ambiguous, acoustic), and treating L1 (English, Cantonese) and Load (intact, English competitor, Cantonese competitor) as fixed factors. There was an effect of L1, $F(1, 7878) = 20.49, p < .001$, but no effect of Load, $F(2, 7878) < 1$, and no L1-by-Load interaction, $F(2, 7878) < 1$. Thus, compared to the English speakers, the Cantonese speakers were relatively less reliant on the lexical-semantic structure of the phrases and comparatively more reliant on the acoustic cues. Importantly, whether the target phrases were intact or suffered from energetic masking, and whether the mask was intelligible to the participants or not had no effect on their relative reliance on acoustic cues vs. lexical-semantic structure.

Lower ratings in Cantonese than English speakers were not simply a by-product of the lower spread in the former than the latter. Indeed, when the two groups were matched on their spread, the main rating difference remained. To match the spread, we rank-ordered the Cantonese participants on the size of their spread, averaged across the three load conditions, and removed as many participants as needed to match the spread of the English group, starting with those with the lowest average spread. We were left with 24 Cantonese participants whose average spread was 3.41 (compared with 2.92 before selection), which was not different from the spread of the English group, i.e., 3.42, $F(1, 4537) < 1$. The analyses described above were run with the reduced Cantonese group and showed exactly the same patterns. Critically, relatively greater reliance on the acoustic cues for the Cantonese than English participants was still manifest, $F(1, 6804) = 13.45, p < .001$, and there was no effect of Load, $F(2, 6804) < 1$, or L1-by-Load interaction, $F(2, 6804) < 1$.

Discussion

Experiment 1 provides answers to the first three questions put forth in the introduction.

1. When the listening environment is devoid of any load, is the balance of weights between lexical-semantic and acoustic segmentation cues similar in native and non-native speakers? No. Within the design and procedure of Experiment 1, the data clearly indicate that non-native speakers give relatively more weight to acoustic cues than to higher-order knowledge. This finding suggests that the non-native speakers' lower familiarity with English allophones (Altenberg, 2005; Ito & Strange, 2009) is not severe enough to critically interfere with speech segmentation. Rather, reliance on acoustic cues appears to be used to compensate for a lack of lexical-semantic knowledge. This is consistent with Cutler et al. (2004, 2008), who showed that speakers' sensitivity to phonetic detail in a second language can be surprisingly high, and likewise, with Mayo et al. (1997), who reported that non-native speakers have greater difficulties exploiting contextual predictability than native speakers. The non-native pattern is also in line with Mattys et al.'s (2005) proposal that reliance on sub-lexical cues prevails when higher-order sources of information are unavailable or incomplete.

2. Are native and non-native speakers' segmentation strategies differentially affected by an energetic load?

No. While the response spread in intact speech was smaller for the non-native than the native speakers, this difference did not increase when the signal was degraded by a competing talker. Likewise, the presence of a competing talker had no effect on the relative weights between lexical-semantic and acoustic cues in either group. Thus, while the effect of background noise

on sentence intelligibility is more detrimental to non-native than to native speakers (e.g., Nábělek & Donahue, 1984), the actual balance of segmentation strategies remains the same in both groups.

3. Does the informational content of a load affect the balance between lexical-semantic and acoustic reliance?

No. In the same way that Mattys et al. (2009) found no difference between intelligible babble noise and speech-shaped noise, Experiment 1 showed no evidence that the intelligibility of the competing talker (L1, L2, or a fully unintelligible language) had any effect on listeners' response spread or cue weights.

Experiment 2

Experiment 2 aimed to answer our fourth question:

4. Does a cognitive load lead to relatively greater reliance on lexical-semantic information in both native speakers (as in Mattys et al., 2009) and non-native speakers?

Mattys et al. (2009) found that native English speakers responded to a cognitive load by relying more heavily on lexical-semantic structure than on acoustic cues (i.e., a lexical drift). In their experiments, a cognitive load was created either by having listeners perform the rating task in one ear and a concurrent task in the other ear, or by having listeners hold verbal materials in short-term memory during the rating task. The lexical drift was taken as evidence that reliance on lexical-semantic information is more robust to a depletion of central processing resources than reliance on sub-lexical information.

Given the non-native speakers' initial bias for acoustically-driven ratings (see intact condition in Experiment 1), we see two possible scenarios for their response to a cognitive load. If the initial non-native bias for salient acoustic cues was due to greater attention to acoustic detail rather than to a lexical-semantic deficit per se, it is likely that the cognitive load will have the same effect in native and non-native speakers, that is, it should cause a lexical drift. However, if the initial non-native bias for salient acoustic cues was due to an inalterable deficit in lexical-semantic knowledge, the cognitive load is unlikely to cause a lexical drift in the non-native speakers.

A cognitive load was created by having listeners perform a visual conjunction search task on an array of colored shapes briefly displayed during the presentation of the test phrases. We used this task because of its non-linguistic nature and non-auditory modality; hence, any effect it might have on the speech task is unlikely to result from simple modality-/domain-specific interference, and more likely to result from depletion of central processing resources (Styles, 1997). After each trial, listeners did the rating task, as in Experiment 1, and then indicated whether or not the visual array contained an item with an incongruent color-shape conjunction.

Method

Participants.

Thirty native English speakers and 30 native Cantonese speakers with no-self-reported hearing problems were recruited for this experiment. They received course credit or a small honorarium for their participation. The English and Cantonese speakers were recruited from the same populations as in Experiment 1. The age range of the Cantonese speakers was 19-22 years. All of them had lived in the UK for a minimum of two years (average: 5 years; range: 2 to 12 years) and had studied English in school since they were eight or under. On a scale from 1 (very poor) to 5 (excellent), they rated their English pronunciation 3.6 on average ($SD = .57$) and their proficiency in English grammar 3.8 ($SD = .58$).

Materials.

The materials for the rating task were the 120 intact phrases of Experiment 1, with 40 of them lexical, 40 ambiguous, and 40 acoustic. The visual arrays used in the cognitive load

condition consisted of 60 grids, each made of seven rows and seven columns. The grids were 9cm by 9cm. The 49 items in each grid consisted of colored shapes. The colors were black, red, yellow, and green. The shapes were squares, diamonds, circles, and triangles. Each grid was made of items of two colors only and two shapes only, which were always combined in the same way, e.g., red circles and black diamonds. These were arranged randomly in the grid (see an example in Figure 2A). Thirty out of the 60 grids contained an "odd-one-out" item. This item had the opposite color-shape combination to the other items in the grid (e.g., a red diamond among red circles and black diamonds, see Figure 2B). The odd-one-out item could appear anywhere in the grid. Following Treisman and Gelade (1980), the number of items in the grid was chosen after piloting sparser grids (e.g., 6 by 6). We settled on a 7-by-7 grid because the pilot data showed accuracy to be around 60 %, ensuring that the task was appropriately demanding (compared to ~ 85 % with a 6-by-6 grid).

Procedure.

For counterbalancing purposes, the 120 phrases were divided into two sets of 60. Each set included 30 phrases containing the target word phrase-initially and 30 phrases phrase-finally. Each subset of 30 phrases contained 15 phrases in which the lexically acceptable parsing arose from the embedding word (e.g., "mild#option") and 15 in which it arose from the embedded word (e.g., "mile#deep"). Within each set of 15 phrases, 5 were lexical, 5 were ambiguous, and 5 were acoustic. The assignment of the two test conditions (intact, visual search) to the two sets of 60 phrases was counterbalanced across pairs of participants in a Latin-square fashion. Thus, each participant was exposed to both test conditions, but never on the same phrases. The two conditions (intact, visual search) were blocked and the two blocks were counterbalanced across participants. Each block was preceded by four practice trials.

The trial-by-trial procedure was the same as in Experiment 1, except that, on the visual-search trials, a grid was briefly displayed in the centre of a computer monitor over a white background during the playback of the phrase. The grid appeared at the beginning of the phrase and disappeared at the end, for an average duration of approximately 1200 ms. On those trials (which were all in the same blocks), participants were asked to pay attention to the grid while listening to the phrase and report whether the grid contained an odd-one-out item after they gave their rating. They did so by pushing one of two buttons labelled "Present" and "Absent" on the computer keyboard. The buttons for the rating task and those for the visual search task were different. The matching between phrases and grids was random. Participants did not know which color/shape combination to expect on any upcoming trial.

Results

Both English and Cantonese speakers performed the visual-search task above 50% chance: 66%, $t(29) = 15.27$, $p < .001$; 60%, $t(29) = 3.48$, $p = .002$, respectively. The English group was marginally better, $F(1, 58) = 3.92$, $p = .05$.

The coding of the ratings was the same as in Experiment 1. Figures 3A and 3B show the average ratings of the lexical, ambiguous, and acoustic phrases in the intact condition and the visual-search condition (labelled "Intact + Visual Search" in figure 3) for the English and Cantonese speakers, respectively. Figure 3C shows the rating spread between the lexical and acoustic phrases. Figure 3D shows the ratings aggregated across the three types of phrases.

As can be seen in Figure 3C, the spread was smaller in the Cantonese group than in the English group. It was also smaller in the visual-search condition than in the intact condition. However, the two language groups were similarly affected by the visual search. A mixed-effect analysis of the spread data, with Phrase (lexical, acoustic), Load (intact, visual search), and L1 (English, Cantonese) as fixed factors, and participants and items as random factors, showed a significant Phrase-by-L1 interaction, $F(1, 4789) = 27.97$, $p < .001$, indicating that the spread

was smaller in the Cantonese than English group. Thus, as already shown in Experiment 1, the non-native speakers showed poorer discrimination between segmentation alternatives than the native speakers. A Phrase-by-Load interaction, $F(1, 4789) = 9.55, p = .002$, indicated that the rating spread was smaller when participants had to simultaneously perform the visual-search task. The three-way Phrase-by-Load-by-L1 interaction was not significant, $F(1, 4789) < 1$, showing that the effect of Load on the spread was comparable in the two language groups.

An analysis with Load (intact, visual search) and L1 (English, Cantonese) as fixed factors, and participants and items as random factors, showed no effect of Load, $F(1, 7181) = 2.12, p = .14$, an effect of L1, $F(1, 7181) = 18.87, p < .001$, and a Load-by-L1 interaction, $F(1, 7181) = 10.53, p = .001$. Thus, as in Experiment 1, Cantonese speakers were relatively more reliant on the acoustic cues than were the English speakers. This was the case in both the intact condition, $F(1, 3589) = 6.68, p < .01$, and the visual-search condition, $F(1, 3692) = 27.67, p < .001$. The visual search caused the English speakers to rely relatively more on lexical structure than acoustic cues, $F(1, 3588) = 12.02, p < .001$. This lexical drift was not observed in the Cantonese group, $F(1, 3593) = 1.34, p = .25$. The difference between the two groups was not the result of a smaller spread for the Cantonese than English speakers. Indeed, when the spread of the Cantonese group was matched with that of the English group, the Cantonese speakers still showed no Load effect. The matching procedure was the same as in Experiment 1, except that the number of Cantonese participants remaining in the analysis after the matching procedure was smaller ($N = 13$). With a spread now comparable to that of the English group (4.25 vs. 4.26), $F(1, 5142) < 1$, the Cantonese group still showed no hint of a lexical drift in response to the visual-search task, $F(1, 1558) = 2.84, p = .09$ —if anything, there was a trend toward an acoustic drift.

Because the Cantonese speakers were slightly less accurate than the English speakers on the visual search task (60% vs 66%), some of the above patterns could be imputed to the two groups having allocated different amounts of processing resources to the segmentation task. However, sub-samples of participants matched on their performance in the visual task showed the exact patterns of segmentation results as above.

Discussion

Experiment 2 highlights three main findings. First, divided attention impaired listeners' capacity to discriminate between segmentation alternatives (as shown by the spread data), even though the speech signal itself did not suffer any energetic masking. This effect, which was already observed in Mattys et al. (2009), confirms the sensitivity of segmentation strategies to a cognitive load. Second, this effect was found despite the fact that the cognitive load was entirely non-linguistic (unlike the cognitive load in Mattys et al., which always had a linguistic component). Thus, the speech-recognition system is penetrable by both linguistic and non-linguistic cognitive loads. Third, in answer to the question posed in the introduction, the cognitive load did not have the same effect on cue reliance in native and non-native speakers. Native speakers responded to the load by giving relatively more weight to lexical information than to acoustic cues, whereas non-native speakers did not. The absence of a lexical drift in the non-native group shows that reliance on acoustic cues remains predominant for non-native speakers even in conditions thought to encourage reliance on lexical-semantic knowledge (Mattys et al., 2009). This result is consistent with the assumption that non-native speakers have severe limitations with lexical-semantic integration and that these limitations manifest themselves as a steady reliance on acoustic cues in intact speech as well as under energetic and informational masking.

We investigated the segmentation strategies used by native and non-native speakers of English in various adverse conditions. Our focus was on conditions involving energetic masking (background competing talker) and on conditions involving solely informational masking (linguistic interference and cognitive load) because these have been shown to produce contrasted response patterns in native speakers (Mattys et al., 2009). We also focused on the relative balance between lexically-driven and acoustically-driven segmentation strategies, which has long been critical in modeling spoken-word recognition (e.g., Davis et al., 2002; Gow & Gordon, 1995; Norris et al. 1995).

Overall, non-native speakers were less successful than native speakers in exploiting the available information to discriminate between segmentation alternatives, and they especially showed low reliance on lexically-driven segmentation. They comparatively relied more strongly on acoustic cues. This pattern probably results from both poor lexical-semantic knowledge and limited integrative capacity—i.e., failure to treat the individual words of a phrase as part of a whole. This would be expected if the time window over which information is integrated is smaller in non-native than native speakers, possibly due to differences in short-term memory capacity (Thorn & Gathercole, 1999). It should be noted that poor lexical-semantic knowledge in non-native speakers could originate as much from a failure to recognize the words making up a phrase as from a blurry distinction between what constitutes a word and what does not. For instance, if non-native speakers were unsure about the lexical status of "doption," they would probably find "*mile#doption" and "mild#option" less lexically contrasted than intended, which would lead to a smaller effect of lexical structure on the ratings.

However, the above pattern does not mean that non-native speakers completely ignore lexical structure to solve ambiguous word boundaries. White, Melhorn, and Mattys (2009) reported clear evidence of segmentation "by lexical subtraction" in Hungarian L2 speakers of English. They showed that a speech fragment (e.g., "corri") was a stronger prime for a subsequent lexical decision task (e.g., printed word "corridor") when preceded by a meaningful word (e.g., "anythingcorri") than by a non-word (e.g., "imoshingcorri"). This effect was reliable even in Hungarian speakers with relatively low English proficiency levels. Thus, non-native speakers do clearly use lexical structure to guide segmentation, but they also show acute sensitivity to acoustic-phonetic cues when lexical structure is not readily extractable.

Interestingly, an acoustic drift comparable to the one exhibited by our Cantonese speakers was reported by Mattys et al. (2009) with native speakers when the amount of energetic masking was very high (22 % glimpsed, compared to 37 % in the present experiment). Incomplete knowledge of a non-native language therefore appears to have processing consequences similar to those experienced by native speakers facing a very incomplete, degraded signal. Thus, whenever inference from higher-order knowledge is challenged, either by reduced experience of that knowledge or by reduced sensory access to it, fallback on salient acoustic-phonetic details is observed.

However, we found no evidence that native and non-native speakers were differentially affected by energetic masking. The two groups showed a comparable reduction in discriminability (cf. spread data) from intact speech to speech in a competing-talker background. This result is inconsistent with the finding that noise is more disruptive to non-native than native listeners (e.g., Mayo et al., 1997; Nábělek & Donahue, 1984). It is more in line Cutler et al. (2004), who found no differential effect of babble noise on the identification of consonant-vowel and vowel-consonant English syllables by native English speakers and Dutch non-native speakers of English. In their study, as in ours, the non-native speakers were less accurate than the native speakers, but the magnitude of the difference was the same in all babble noise conditions. Cutler et al. (2008) subsequently showed that the lack of interaction between

L1 and noise, compared to the standard pattern evident in a conceptually similar study by Garcia Lecumberri and Cooke (2006), was the result of seemingly small task and material differences, e.g., babble noise leading time, blocking of conditions, variability of syllable components. They argued that these differences were critical in allowing (or not) native speakers to use cues from the design context to compensate for masking. Similarly, the absence of an interaction between L1 and Load in the spread data of Experiment 1 could be the result of both high reliance on lexical-semantic information by native speakers in the intact condition and high reliance on acoustic cues by non-native speakers in the competing-talker condition. In either case, our results confirm that disproportionately poorer performance for non-native listening in noise is far from being a ubiquitous phenomenon.

Furthermore, the presence of the competing talker had no effect on the balance of segmentation strategies in either L1 group. In particular, it did not cause non-native speakers to fall back on either acoustic cues or lexical-semantic structure. A more detailed inspection of the data for the non-native group (Figure 1B) suggests an interesting pattern, however. While the presence of a competing talker did not have a significant effect on the ratings of the lexical phrases ($F[2, 1314] = 2.33, p = 10$), it caused the ratings of the acoustic phrases to be strongly biased towards the lexically acceptable alternative ($F[2, 1312] = 12.17, p < .001$). Similar comparisons in the native group (Figure 1A) revealed significant drifts for both types of phrases, $F(2, 1312) = 9.48, p < .001$, $F(2, 1316) = 12.39, p < .001$, respectively. Thus, from these analyses, it appears that non-native speakers' reliance on lexical-semantic structure was relatively spared by the competing talker compared to their reliance on acoustic cues. This conclusion should be taken with caution, however, because it could be argued that the flat pattern in the lexical phrases for the non-native speakers was more the result of low reliance on lexical structure in the intact condition than genuine tolerance to a competing background voice. Further evidence for (or against) differential weighting of segmentation strategies in noise by native and non-native speakers is needed.

The language of the competing talker (English vs. Cantonese) was inconsequential. Specifically, the competing Cantonese talker did not impact on the Cantonese speakers' performance more than did the competing English talker (and vice versa), and it did not lead to any reweighting of cues. Imperviousness to the content of a competing utterance was also observed by Mattys et al. (2009), who found no difference between the effect of an intelligible competing talker and the effect of its speech-shaped equivalent. This result is inconsistent with the studies by Van Engen and Bradlow (2007) and Garcia Lecumberri and Cooke (2006), in which a distracting voice in a native language was found to be more detrimental to sentence and phoneme intelligibility than a distracting voice in an unknown language—note that, in the Garcia Lecumberri and Cooke study, the effect was small and asymmetrical. A number of stimulus and design differences between the Van Engen and Bradlow study and ours were reviewed in Mattys et al., including dissimilarities and numbers of babble talkers, as well as methods for matching energetic masking across babble conditions. However, perhaps most important is the fact that our task was geared very specifically to speech segmentation, independent of intelligibility per se. Our phrases were structurally repetitive, our response sets were small (e.g., "mild" vs. "mile") and made available to the listeners, and the target phrases were short. In the Van Engen and Bradlow (2007) study, which showed the strongest evidence of linguistic interference, the stimuli were meaningful sentences and the task was open-ended transcription. The design features of our experiments could have helped listeners maintain a high level of attention to task-relevant characteristics of the phrases and segregate the two streams effectively, keeping interference to a minimum. This could explain why, beyond reducing discriminability due to energetic masking, our competing talkers had no informational

effect on the segmentation task. One way to test the relationship between stream segregation and language interference would be to use competing utterances less conducive to segregation, e.g., utterances exhibiting a greater spectral overlap with the target phrases—although this would automatically increase energetic masking as well—or utterances showing semantic associations with the target phrases. We suspect that language interference would be more likely to appear in those conditions. A different, though not incompatible, possibility is that our non-native listeners were good at segregating the target phrases from the competing utterances because of their extended experience with English. Less proficient speakers might be less capable of fending off attentional capture from their native language.

As for the effect of a purely informational (i.e., cognitive) load on speech segmentation, native and non-native speakers showed a similar loss of discriminability when asked to perform a visual search task while listening to the test phrases. However, their balance of segmentation strategies in response to the load differed markedly. The native speakers relied relatively more strongly on lexical-semantic than on acoustic cues. This lexical drift was already observed by Mattys et al. (2009) with a cognitive load involving linguistic materials, e.g., divided attention between two speech messages played dichotically or holding words in short-term memory during the playback of the test phrases. The present result shows that greater attention to lexical-semantic plausibility (and/or decreased attention to acoustic details) is a widespread response to loads taxing central processing resources, at least among native speakers. The fragility of acoustic perception under a cognitive load in native speakers was also illustrated by Casini, Burle, and Nguyen (2009), who found that native French speakers' discrimination of CVC words differing solely by the duration of their vowel was significantly biased towards the word with the shorter vowel when listeners' attention was divided between the word identification task and a concurrent arithmetic task. However, the exact locus of the lexical drift under cognitive load remains to be established. At issue is whether it originates from a reduction in auditory acuity (locus = phonetic perception) or from an emphasis on communicatively valid responses (locus = lexical and post-lexical processing).

The non-native speakers, in contrast, did not show any sign of a lexical drift; their initial strong reliance on acoustic cues (cf. the intact condition) was unaffected by the cognitive load. This result highlights non-native speakers' limited ability to exploit lexical-semantic information to compensate for input ambiguity. Research on the effect of sentential context on the disambiguation of homophones has revealed similar limitations in non-native speakers compared to native speakers (Elston-Güttler & Friederici, 2005; Love, Maas, & Swinney, 2003). Thus, we argue that the effect of a cognitive load on segmentation is a function of listeners' intrinsic weights for lexical-semantic and acoustic cues, with high-weight cues more tolerant to a depletion of processing resources than low-weight cues. In the native English group, this would indeed translate into reduced reliance on acoustic cues, which are thought to be outweighed by lexical information (Mattys et al., 2005). In the non-native group, the relatively higher weight of acoustic cues compared to the native group would make acoustic cues comparatively more resilient to a cognitive load. Although we did not observe a significant acoustic drift in that group, which, strictly speaking, would be a logical correlate, the absence of a lexical drift is probably a compromise between high reliance on acoustic cues and the need (albeit unfulfilled) for communicatively viable segmentation.

In summary, we found significant differences between native and non-native segmentation strategies in intact speech, as well as under adverse listening conditions. In all conditions, non-native speakers were consistently less successful in discriminating segmentation alternatives, and they tended to overemphasize salient acoustic cues. However, they were not more adversely affected by energetic masking (competing background talker) than were the native speakers.

Therefore, the claim that speech recognition in noise is more challenging to non-native than native speakers seems to critically depend upon the task in which listeners are engaged and the type of resources made available to them. In the present segmentation task, the signal was apparently rich enough for the non-native speakers to use adequate compensation mechanisms. Likewise, non-native speakers were not more likely than native speakers to be distracted by the intelligibility of the competing utterances, possibly because of non-native speakers' greater focus on acoustic cues. This bias was also manifest under a cognitive load: While a cognitive load led to decreased reliance on acoustic detail in the native speakers, it did not in the non-native speakers.

References

- Altenberg, E. P. (2005). The perception of word boundaries in a second language. *Journal of Second Language Research*, 21, 325-358.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subject and items. *Journal of Memory and Language*, 59, 390-412.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (Release 2)* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].
- Barker, J. & Cooke, M. P. (2007). Modelling speaker intelligibility in noise. *Speech Communication*, 49, 402-417.
- Bradlow, A. R. & Alexander, J. A. (2007). Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 121, 2339-2349.
- Bradlow, A. R. & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, 112, 272-284.
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, 109, 1101-1109.
- Casini, L., Burle, B., Nguyen, N. (2009). Speech perception engages a general timer: Evidence from a divided attention word identification task. *Cognition*, 112, 318-322.
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America*, 119, 1562-1573.
- Cooke, M. P., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail effect party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America*, 123, 414-427.
- Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. P. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *Journal of the Acoustical Society of America*, 124, 1264-1268.
- Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America*, 116, 3668-3678.
- Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 218-244.
- Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24, 423-444.
- Elston-Güttler, K. E. & Friederici, A. D. (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory & Language*, 52, 256-283.
- Farris, C., Trofimovich, P., Segalowitz, N., & Gatbonton, E. (2008). Air traffic communication in a second language: Implications of cognitive factors for training and assessment. *TESOL Quarterly*, 42, 397-410.
- Festen, J. M. & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, 88, 1725-1736.
- Garcia Lecumberri, M. L. & Cooke, M. P. (2006). Effect of masker type on native and non-native consonant perception in noise. *Journal of the Acoustical Society of America*, 119, 2445-2454.
- Gow, D. W. & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception*

and Performance, 21, 344-359.

Hazan, V. & Simpson, A. (2000). The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects. *Language and Speech*, 43, 273-294.

Ito, K. & Strange, W. (2009). Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *Journal of the Acoustical Society of America*, 125, 2348-2360.

Love, T., Maas, E., & Swinney, D. (2003). The influence of language exposure on lexical and syntactic language processing. *Experimental Psychology*, 50, 204-216.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134, 477-500.

Mayo, L. H., Florentine, M., & Buss, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, 40, 686-693.

Nábělek, A. K. & Donahue, A.M. (1984). Perception of consonants in reverberation by native and non-native listeners. *Journal of the Acoustical Society of America*, 75, 632-634.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1209-1228.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, 54, 1235-1247.

Rosenhouse, J., Haik, L., & Kishon-Rabin, L. (2006). Speech perception in adverse listening conditions in Arabic-Hebrew bilinguals. *International Journal of Bilingualism*, 10, 119-135.

Sanders, L. D. & Neville, H. J. (2003). An ERP study of continuous speech processing: II. Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research*, 15, 214-227.

Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research*, 45, 519-530.

Styles, E. A. (1997). *The psychology of attention*. Bucks, UK: Taylor and Francis Psychology Press.

Takano, Y. & Noda, A. (1993). A temporary decline in thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology*, 24, 445-464.

Thorn, A. S. C. & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology*, 52, 303-324.

Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.

Van Engen, K. J. & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America*, 121, 519-526.

Van Wijngaarden, S., Steeneken, H., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America*, 111, 1906-1916.

White, L., Melhorn, J. F., & Mattys, S. L. (2009, in press). Segmentation by lexical subtraction: Hungarian bilinguals listening to English speech. *Quarterly Journal of Experimental Psychology*.

Authors' Note

Correspondence can be sent to Sven Mattys, Department of Experimental Psychology, University of Bristol, 12A Priory Road, Bristol BS8 1TU, UK, E-mail: Sven.Mattys@bris.ac.uk. This study was made possible thanks to a grant from the Leverhulme Trust (F/00 182/BG) to S. L. Mattys, and a Research Training Network grant from the Marie Curie foundation (MRTN-CT-2006-035561). We thank Martin Cooke for calibrating the babble noise and calculating the glimpsing percentages. We also thank Lukas Wiget for contributing to data collection and Jeff Bowers for comments on an earlier draft.

Figure Captions

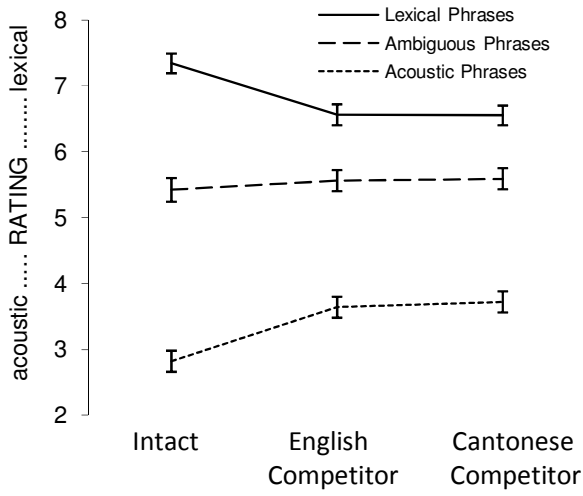
Figure 1. Mean ratings (and error bars) of lexical, ambiguous, and acoustic phrases in intact speech, in a competing English-talker background, and in a competing Cantonese-talker background for native English speakers (1A) and for native Cantonese speakers of second-language English (1B). Figure 1C shows the difference between lexical and acoustic phrases (spread) for the two language groups. Figure 1D shows ratings aggregated across the lexical, ambiguous, and acoustic phrases for the two language groups.

Figure 2. Examples of visual displays used as cognitive load in Experiment 2. Figure 2A shows a target-absent display. Figure 2B shows a target-present display, with the odd-one-out item (a red diamond) in the third column and second row.

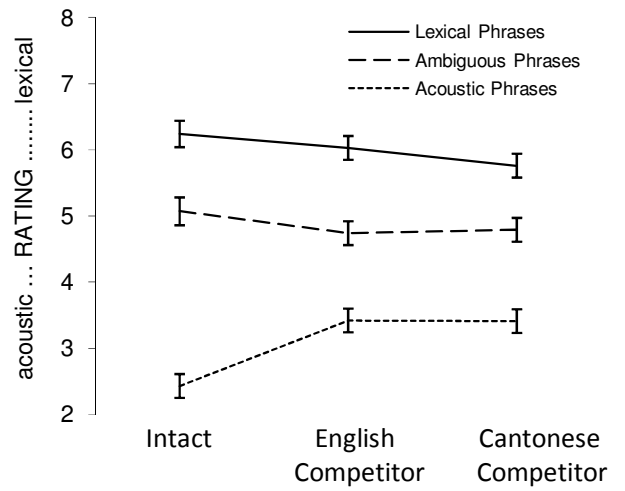
Figure 3. Mean ratings (and error bars) of lexical, ambiguous, and acoustic phrases in intact speech without and with a concurrent visual search task for native English speakers (3A) and for native Cantonese speakers of second-language English (3B). Figure 3C shows the difference between lexical and acoustic phrases (spread) for the two language groups. Figure 3D shows ratings aggregated across the lexical, ambiguous, and acoustic phrases for the two language groups.

Figure 1.

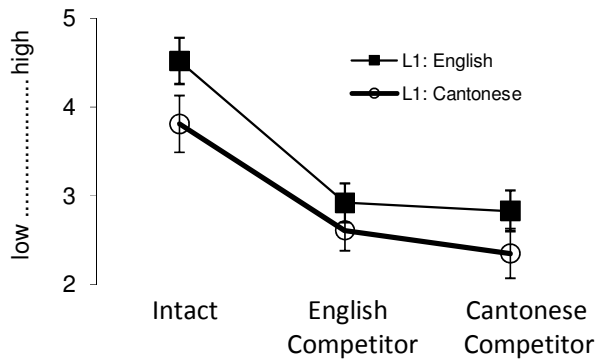
A. Listeners' L1: English



B. Listeners' L1: Cantonese



C. Spread



D. Aggregated Ratings

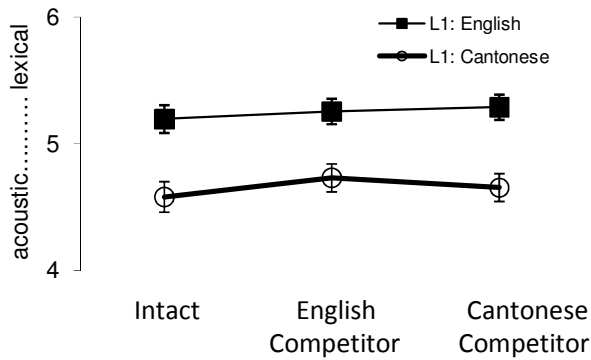
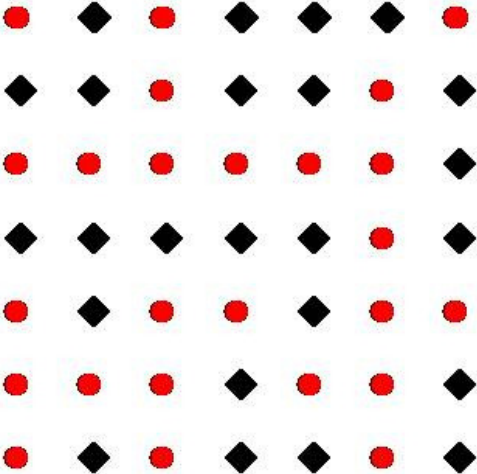


Figure 2.

A



B

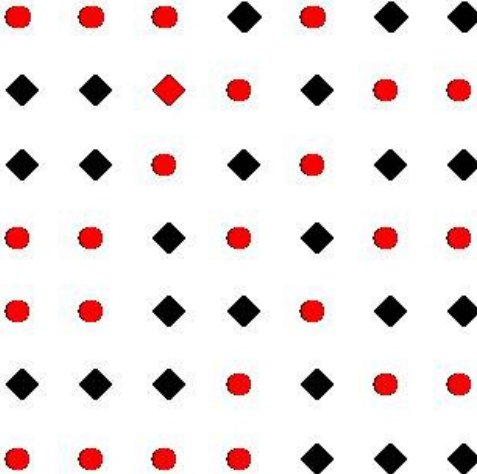
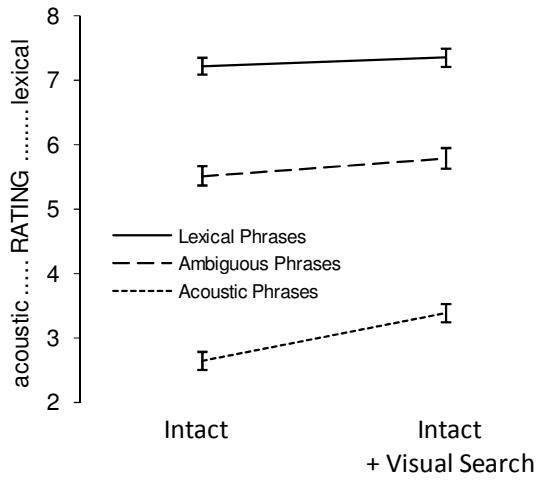
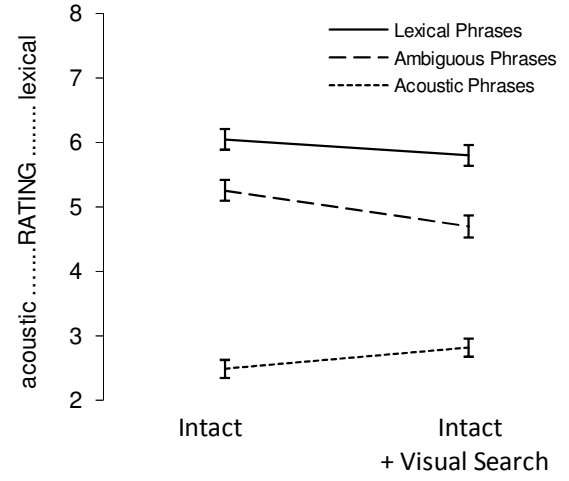


Figure 3.

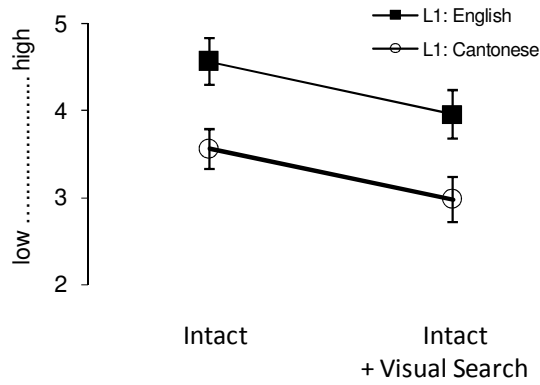
A. Listeners' L1: English



B. Listeners' L1: Cantonese



C. Spread



D. Aggregated Ratings

