



**HAL**  
open science

# Influence of English r-resonances on intelligibility of speech in noise for native English and German listeners

Antje Heinrich, Yvonne Flory, Sarah Hawkins

► **To cite this version:**

Antje Heinrich, Yvonne Flory, Sarah Hawkins. Influence of English r-resonances on intelligibility of speech in noise for native English and German listeners. *Speech Communication*, 2010, 52 (11-12), pp.1038. 10.1016/j.specom.2010.09.009 . hal-00698850

**HAL Id: hal-00698850**

**<https://hal.science/hal-00698850>**

Submitted on 18 May 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

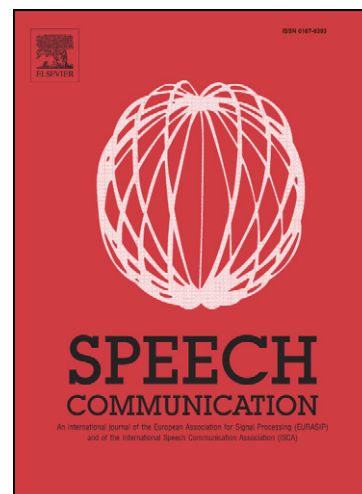
Influence of English r-resonances on intelligibility of speech in noise for native English and German listeners

Antje Heinrich, Yvonne Flory, Sarah Hawkins

PII: S0167-6393(10)00159-7  
DOI: [10.1016/j.specom.2010.09.009](https://doi.org/10.1016/j.specom.2010.09.009)  
Reference: SPECOM 1939

To appear in: *Speech Communication*

Received Date: 16 November 2009  
Revised Date: 29 September 2010  
Accepted Date: 30 September 2010



Please cite this article as: Heinrich, A., Flory, Y., Hawkins, S., Influence of English r-resonances on intelligibility of speech in noise for native English and German listeners, *Speech Communication* (2010), doi: [10.1016/j.specom.2010.09.009](https://doi.org/10.1016/j.specom.2010.09.009)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Influence of English r-resonances on intelligibility of speech in noise for  
native English and German listeners**

Antje Heinrich<sup>1</sup>

Yvonne Flory<sup>2</sup>

Sarah Hawkins<sup>1</sup>

1 Department of Linguistics, University of Cambridge, UK

2 Allgemeine Linguistik, Universität des Saarlandes, Saarbrücken, Germany

Address correspondence to:

Antje Heinrich

University of Cambridge, Department of Linguistics

Sidgwick Avenue

Cambridge, U.K. CB3 9DA

Telephone: +44 1223 760814

Fax: +44 1223 335053

e-mail: [ah540@cam.ac.uk](mailto:ah540@cam.ac.uk)

**ABSTRACT**

Non-rhotic British English speakers and Germans living in England were compared in their use of short- and long-domain **r**-resonances (cues to an upcoming [ɹ]) in read English sentences heard in noise. The sentences comprised 52 pairs differing only in /r/ or /l/ in a minimal-pair target word (*mirror, miller*). Target words were cross-spliced into a different utterance of the same sentence-base (match) and into a base originally containing the other target word (mismatch), making a four-stimulus set for each sentence-pair. Intelligibility of target and some preceding unspliced words was measured. English listeners were strongly influenced by **r**-resonances in the sonorant immediately preceding the critical /r/. A median split of the German group showed that those who had lived in southeast England for 3-20 months used the weaker long-domain **r**-resonances, whereas Germans who had lived in England for 21-105 months ignored all **r**-resonances, possibly in favour of word frequency. A preliminary study of German speech showed differences in temporal extent and spectral balance (frequency of F3 and higher formants) between English and German **r**-resonances. The perception and production studies together suggest sophisticated application of exposure-induced changes in acoustic-phonetic and phonological knowledge of L1 to a partially similar sound in L2.

## 1. Introduction

Natural speech, especially in the context of connected discourse, often retains a high degree of intelligibility even in challenging listening situations such as background noise (Bronkhorst, 2000 and references cited earlier). This observation may be explained, at least in part, by the wide range of pragmatic, semantic, structural-linguistic and acoustic-phonetic properties of natural speech which together provide a high degree of signal redundancy: if one type of information is missing, hard to hear, or unexpected, listeners can often use another type to understand the message.

One way to conceptualise these multiple influences on intelligibility is in terms of coherence of multiple types of information. When different types of information are mutually compatible and thus converge on a particular interpretation of the signal, then that interpretation is likely to win over others. The term ‘perceptual coherence’ and others like it have been used independently by a number of people (e.g., Grunke & Pisoni, 1982; Hawkins, 1995, 1996, 2003; Remez, 1994, 2003; Remez, Rubin, Berns, Pardo, & Lang, 1994) to express the sense that, for a signal to be understood, it must ‘hang together’ and sound as if it comes from one speaker. Hawkins (2003, p. 384) suggested that the concept of coherence “is part hypothesis, part factually based: we do not know exactly what properties make speech perceptually coherent, but we do know from many different types of work that small perturbations can change its perceived coherence (Darwin & Gardner, 1985; Huggins, 1972a, 1972b). It is rooted in the sensory signal but relies on knowledge; the two are not distinct in this respect, but feed each other.”

The interplay between different types of knowledge and physical properties of the acoustic signal is the focus of this paper. The particular emphasis is on acoustic-phonetic coherence of the physical signal in indicating the presence or absence of an upcoming /r/ phoneme, and

how that physical information is used by native and non-native speakers of the language to identify words in read sentences heard in background noise.

The choice of /r/ over other phonemes is deliberate, because /r/ is unusual—and probably unique—amongst English phonemes in the duration over which traces of its articulation can be found, and thus in having a long-term distinctive acoustic signature of its own, called *r*-resonance (Kelly & Local, 1986). Other properties of sounds do of course spread over several segments, the best-known example being lip rounding (Benguerel & Cowan, 1974), and others being coda voicing (Coleman, 2003; Hawkins & Nguyen, 2004) and the feature [anterior] (Coleman, 2003) which in English distinguishes alveolars /s z t d/ from postalveolars /ʃ ʒ tʃ dʒ/. However, these properties are better analysed as distinctive (phonological) features which act as cues to a group of phonemes (sometimes in particular structural positions) rather than to a single phoneme. English *r*-resonances, on the other hand, seem to be associated with just that phoneme (Kelly & Local, 1986). Moreover, whereas lip rounding spreads for perhaps 200 ms (Bell-Berti & Harris, 1982; Coleman, 2003) over several segments with which it happens to be compatible, and acoustic correlates of coda voicing probably spread at most only to a preceding unstressed syllable (Coleman, 2003), *r*-resonances in Standard Southern British English (SSBE) and several other non-rhotic British accents spread over longer durations, commonly up to about 600 ms, with one instance of 1000 ms on record (Coleman, 2003; Heid & Hawkins, 2000; Kelly & Local, 1986; West, 1999a, 1999b). (The term *non-rhotic* refers to accents in which an orthographic postvocalic /r/ is pronounced only when also followed by a vowel, thus /maɪ mɪr/ for *my mirror* and /draɪv ðə kɑːr ɒvər hi/ for *drive the car over here*, but /maɪ kɑːz ɒvər hi/ for *my car's over here*. Unless specifically mentioned, all English speech discussed in this paper refers to non-rhotic British accents.)

Figure 1 shows data from Heid and Hawkins (2000), who measured formant frequencies of vowels in carrier phrases, 5-7 words long, that contained just one /l/ or /r/, always in the final word e.g. *we heard that it might be a mirror/miller*. They found a pronounced decrease in F3 and F4 frequency for the directly preceding vowel when it was followed by an r-word compared to an l-word. They also showed a smaller yet long-lasting decrease in F3 and F4 for most vowels up to about 600 ms before the critical approximant. Changes in F2 were smaller and somewhat less consistent across contexts. As discussed by West (1999b) and by Heid and Hawkins (2000), the magnitude (and in a few cases the direction) of differences in formant frequency vary with speaker, vowel quality, stress and segmental context. Measurements consistent with Heid and Hawkins' (2000) data are reported by Tunley (1999), Hawkins and Slater (1994) and Coleman (2003).

INSERT FIGURE 1 ABOUT HERE

Although the measured differences between vowels with and without long-domain r-resonances are small, they can have a surprisingly strong influence on native English speakers' perceptual decisions, at least in adverse listening conditions. In a two-alternative forced choice task using natural speech, West (1999b) showed that native English listeners could tell which one of a minimal pair contrasting in the presence of a word-medial /r/ or /l/ e.g. *mirror* or *milller*, had been spoken in the carrier phrase *No, I uttere(d) \_\_\_\_ today*, when the [ɹ] or [l] itself—and often more of the signal—was replaced by noise. Hawkins and colleagues (Hawkins & Slater, 1994; Tunley, 1999) found similar results using synthesized carrier sentences with or without r-resonances, in a free-response word-identification task in which the words could be real or nonsense. Heard in cafeteria noise, the long-term resonances not only informed about the presence of the /r/ sound in the target word, but also increased intelligibility of other words in the utterance.

The above experiments attest to the perceptual relevance of **r**-resonances when the choices are either very restricted (West, 1999b) or the signal is otherwise impoverished by virtue of being synthesized (Hawkins & Slater, 1994; Tunley, 1999). They do not tell us whether **r**-resonances influence perception of more natural speech. The present study examines the role of **r**-resonance on word intelligibility when the speech signal comprises meaningful, naturally-read sentences and the choice of words is unrestricted. As just explained, this question is of interest in itself for native speakers of many accents of British English. By including listeners whose native language is German, we could also begin to assess the role of physical coherence of the signal in interaction with phonetic and other linguistic knowledge.

A number of reasons motivated the choice of German, some of them independent of the purposes of the present paper. However, one pertinent reason is that though the main place of oral constriction is alveolar for SSBE /r/ and uvular for standard German /r/, both are readily heard as performing the phonological function of /r/ by native speakers of the other language, even if they cannot make the correct sound themselves. This is not always the case for other varieties of /r/ in other pairs of languages (and even other dialects of English). For example, Ladefoged (2001) notes that the alveolar tap [ɾ] functions as /r/ in most Scottish accents but (intervocally) as /t/ in American English, such that the American pronunciation of *petal* with a voiced alveolar tap sounds like *pearl* to someone from Edinburgh, Scotland, the stressed vowels also being similar in quality. That is, in Lindau's (1985) Wittgensteinian terms, German and SSBE /r/s have a closer 'family resemblance' than many other pairs of rhotics. These properties of the two /r/ sounds mean that German native speakers listening to English as an L2 have an interesting challenge: they readily hear the English /r/ as functionally equivalent to their native German /r/, and some of the spectral properties are similar; but crucially, others differ markedly.



The standard German syllable-onset /r/ is a uvular fricative [ʁ] (e.g., Ulbricht, 1972) although variants include trills and approximants (Lodge, 2003; Simpson, 1998). The lips are not characteristically rounded, and excitation may be glottal (periodic or aperiodic) and/or uvular (fricated), some of these varieties being context-dependent allophones (Simpson, 1998). The standard English /r/ under discussion in this paper is a postalveolar approximant [ɹ] in which the lips are normally somewhat rounded or protruded, and the back of the tongue is bunched or retracted to a degree which is both accent-specific and subject to significant individual variation. Excitation is normally glottal (periodic or aperiodic) though frication may be observed in some allophonic contexts, especially with other coronal articulations as, for example, in the onset of the word *treat*. Ogden (2009) gives an excellent comprehensive yet brief description of the complex phonetic and phonological properties of this sound in English. Laver (1994) offers a fuller phonetic description of some of its aspects. Experimental investigations spanning many years, usually of American [ɹ], concur that there are many different ways to produce auditorily similar sounds; for recent examples that include MRI see Nieto-Castanon et al. (2005) and Zhou et al. (2008); for a comprehensive and well illustrated yet succinct discussion of American /r/ see Secord et al. (2007), Chapter 5.

Acoustically, English [ɹ] is characterised by a slightly raised F1 relative to other approximants (including its companion liquid, [l]), and very low frequencies in the higher formants, which have very low amplitudes when the oral constriction is narrowest. The formants whose frequencies fall the most are mainly determined by the phonetic quality of the surrounding sounds (so in a VCV context the most affected formants in the [ɹ] segment are strongly vowel-dependent), but in general, F3 and/or F4 frequencies fall dramatically; F2 is usually less affected, and F5 normally follows the same type of trajectory as F4 (e.g., Stevens, 1998). The spectral centre of gravity of English [ɹ] is thus skewed towards the lower frequencies,

with some or all of the higher formants markedly low in frequency and in amplitude, and often closely bunched together. In contrast, formant frequencies of German [ʀ] are more evenly spread across the spectrum. F1 (present in the phonated but not the fricated variant) is not especially low, but F2 is somewhat low (compared, say, to schwa) while F3 is high and often close to a high F4; all formants have fairly high amplitude (Machelett, 1996; Simpson, 1998).

The significance of these differences in segmental articulation of German and English /r/ to the issue of long-domain **r**-resonances is as follows. Though German and English /r/ differ in primary place of articulation, they have in common that the tongue is backed. However, while tongue backing is a primary and thus obligatory part of the German uvular articulation, it can be only a secondary articulation for the English postalveolar approximant. Moreover, it has been suggested that tongue backing may be the basis of long-domain **r**-resonance in English because it may reflect the acoustic consequences of anticipatory (and perseverative) velarisation or pharyngealisation (e.g., West, 1999a). We do not distinguish velarisation from pharyngealisation hereafter, using velarisation for both. Since tongue backing is both obligatory for German /r/ and also appears to colour at least some neighbouring segments (Barry, 1995; Lodge, 2003) German /r/ thus forms a promising base for comparison with English **r**-resonances, a point that is echoed by Lodge (2003), who suggests that word-final German /r/ and non-rhotic British English word-initial and word-medial /r/ have very similar effects on the immediately surrounding segments. On the other hand, it is not clear to what extent English and German will show similar temporal spreads of **r**-resonance effects given that tongue backing is essential to achieve the primary constriction for [ʀ], but performs a more secondary function for [ɹ]. Since we know of no published measurements of long-domain **r**-resonances in German, we therefore conducted a small-scale study to assess whether German [ʀ] is accompanied by long-domain resonances.

## 2. Experiment 1: preliminary acoustic measures of German r-resonances

### 2.1. Participants

Participants were ten native speakers of German, five men and five women, mean age 25.5 years (range 24 - 30). All volunteers reported normal speech and hearing and were students of the Universität des Saarlandes in Saarbrücken. Speakers originated from different regions within Germany, but had been residents of Saarbrücken for at least the last 3.5 years. Five participants spoke standard German without a regional accent, two spoke with a Saarbrücken regional accent, and three had a Swabian accent. All had uvular fricative /r/s: [ʀ].

### 2.2. Methods

#### 2.2.1. Materials

Critical words were 18 minimal pairs differing in the presence of a single /l/ vs /r/. In nine of these pairs the /r/ or /l/ was word-initial (e.g. *Lüge/Rüge*), in six it was in an initial consonant cluster (e.g. *Gleise/Greise*) and in three pairs it was word-medial (e.g. *Kehle/Kehre*). The complete list is in Appendix 1. Each critical word was placed in one of two carrier phrases depending on whether the target word was singular or plural. Neither carrier phrase contained an /r/ or /l/. Thus the members of each sentence-pair differed in only one phoneme, /r/ or /l/, and this difference was always in the critical word. For example: 1) Singular: *Ich habe heute eine Lüge/Rüge gesehen*. Plural: *Ich habe heute einige Gleise/Greise gesehen*. Words containing an /l/ were chosen as contrasts for words containing an /r/ in order to ensure comparability with acoustic measurements of r-resonances in English speech (Heid & Hawkins, 2000; West, 1999b).

### 2.2.2. *Recording*

The sentences were presented in a list, which each speaker read twice in a quiet room at the Universität des Saarlandes using an AKG-Headset C420III PP with condenser microphone. Recordings were made using Multispeech software and an Audigy 2 ZS Creative sound card with a sampling frequency of 44.1 kHz.

### 2.2.3. *Acoustic measurements*

One token of each word was measured. The beginning and end of the periodic portion corresponding to the word-final schwa in each *heute* and *eine/einige* were marked manually using standard segmentation criteria, and the frequencies of the first four formants were measured at the midpoint of each schwa. Formant frequencies were measured automatically in Praat using a 30-ms Burg lpc spectrum with a Gaussian-like window. The formant search range was adjusted individually for each speaker to yield estimates for four formants. (This is the equivalent, in Praat, of adjusting numbers of poles in the lpc spectrum.) For male speakers, the maximum frequency of the search range was set between 4500 and 5000 Hz, for female speakers between 5000 and 5500. All values computed in this way were checked against a spectrogram of the utterance and manually measured if necessary. The data were analysed using paired-sample t-tests for each formant separately.

In addition, the duration between the midpoint of each schwa and the onset of the acoustic segment corresponding most closely to the following /r/ or /l/ was measured.

### 2.3. *Results*

Tables 1A, 1B and 1C show the frequencies of the first four formants (and standard deviations) of the schwas in *heute*, *eine* and *einige*, for, respectively, the critical /r/ and /l/ in

singleton word-initial, clustered word-initial, and word-medial positions. Details of significance tests are in Table 1D.

INSERT TABLE 1 (A-D) ABOUT HERE

When the /r/ or /l/ is word-medial, there are no significant differences in any of the schwas in the preceding words (Table 1C). There is, however, an interesting pattern of differences when the /r/ or /l/ is in the beginning of the word, whether as a singleton or in a cluster (Tables 1A and 1B respectively). The most consistent aspect of this pattern is that F2 lowers in the /r/ context in the schwas of *eine* and *einige* (Tables 1A and 1B). This F2 lowering may extend as far as the schwa in *heute* when the conditioning /r/ is not in a cluster and is preceded by *eine* e.g. *ich habe heute eine Rüge gesehen* (Table 1A). Consistent with expectations for uvulars (Simpson, 1998), F1 is higher in /r/ contexts than in /l/ contexts in the schwa immediately adjacent to word-initial unclustered /r/ (*eine* and *einige*) and in *heute* before *eine* (Table 1A). Again as expected for uvulars, F3 remains high in all cases (Tables 1A-C); perhaps in consequence of this, F3 shows only 2/12 significant differences across all cases, both of them in the schwa immediately adjacent to the target /r/ or /l/ in the word-initial singleton condition (Table 1A). Finally, F4 is lower in the schwa of *eine* before word-initial singleton /r/ than /l/ (Table 1A).

For *einige*, the same basic pattern holds for F1, F2 and F4, but significance levels are weaker (Table 1D). Moreover, where there is a difference, F3 is lower rather than higher in /r/ contexts. These patterns are almost certainly caused by the velar consonant in *einige*, as confirmed by the absence of significant differences for *einige*'s schwa in the initial cluster cases (Table 1B), where two out of the three critical words also begin with a velar (*Gleise/Greiser*, *Gläser/Gräser*). In contexts other than high front vowels, velars engender a low F3 close to F2 (Stevens, 1998). Heid and Hawkins (2000) used the same reasoning to

explain the smaller measurable resonance effects on *could*, with its velar onset /k/ and back vowel, than on *might*, with its labial onset and front vowel (Figure 1).

#### 2.4. Discussion

These preliminary data allow us to tentatively conclude that the standard German /r/, the uvular fricative [ʀ], shows weak signs of long-domain resonances which are quite strongly context-dependent. The patterns that are present are as expected given the acoustic properties of the segment (Machelett, 1996; Simpson, 1998): relative to the influence of clear [l], German uvular /r/ lowers F2 and to a lesser extent F4 frequencies in an adjacent preceding schwa, while the uvular's high F1 and F3 are likewise anticipated. As expected, the influences are strongest when the schwa is immediately adjacent to the critical /r/ segment (*eine* and *einige* before a word-initial /r/. Table 1A); the presence of another consonant in the word onset is enough to restrict the reliable shifts in formant frequency to F2 only (Table 1B).

The temporal spread is not great: never more than 300 ms in these data. Indeed, it is usually of the order of 100 ms and only approaches 300 ms in one case, that of *heute* when the critical segment is word initial and not in a consonant cluster (Table 1A) and even then, only when the next word is *eine*. The absence of r-resonances in *heute* before *einige* cannot be explained by a greater difference in duration between *heute* and the critical phoneme in this context. Table 1A shows that the duration between *heute* and the critical phoneme is not shorter in the *eine* than the *einige* context, despite *einige* having one more syllable. As noted above, the English data of Heid and Hawkins (2000) suggest that constraints imposed by intervening consonants may exert an influence on the spread of r-resonances, so in these data the spread of any F1 and F3 resonance effects might be 'blocked' by the /r/ of *einige*. This issue is not easily addressed due to the strong constraints that German grammar imposes on sentence construction; in any case, it is a question for future work. Here, we conclude that these results

are quite different from the typical 300-600 ms range for English as cited in the literature discussed earlier and that there is evidence for limited spread of weak **r**-resonance in German, in restricted contexts. Importantly for the main focus of this paper, the strongest and most widespread difference between /r/ and /l/ contexts in German is in F2 frequency, whereas for English it is in F3 and F4, with less consistent long-domain effects on F2 frequency (Heid & Hawkins, 2000; West, 1999b).

### **3. Experiment 2: Intelligibility of English sentences by L1 English and L2 German listeners**

This experiment examined the role of **r**-resonances on word intelligibility of meaningful, naturally-read sentences with an unrestricted choice of words. Both native English and native German listeners were tested on English sentences that were presented in a background of 12-talker babble. A substantial body of literature indicates that perceiving speech in noise is much more difficult for non-native than native listeners (e.g., Bradlow & Alexander, 2007; Cutler, Lecumberri, & Cooke, 2008; Heinrich, Schneider, & Craik, 2008; Lecumberri & Cooke, 2006; Mayo, Florentine, & Buus, 1997), and that non-native listeners require a better signal-to-noise ratio (SNR) in order to achieve similar intelligibility as native speakers (Bradlow & Alexander, 2007; Mayo et al., 1997). Therefore, German listeners heard the stimuli at a more favourable SNR (by 3 dB) than native English listeners. This value of 3 dB was chosen because Heinrich et al. (2008) found in a similar listening situation that it effectively equated intelligibility scores between native and highly proficient non-native listening groups. Importantly, all L2 speakers in the listening experiment (Experiment 2) were proficient users of English who had lived in an English-speaking country for at least 5 months and in Southern England for at least the last 3 months. This criterion ensured fairly lengthy and recent exposure to **r**-resonances in Southern Standard British English (SSBE). Most

listeners had much longer exposure to native spoken English (see Table 3). All had excellent English vocabulary.

The particular phonemic categories being manipulated (/r, l/) behave similarly in phonological terms, and, as shown above, have partially similar phonetic properties in English and German. Two points are at issue. The first is whether long-domain **r**-resonances are perceptually salient for L1 listeners under adverse listening conditions when grammatical forms and word choices are unrestricted. The second is whether L2 listeners use short- and/or long-domain English **r**-resonances under such natural conditions.

In addition to the many variables introduced by the use of unrestricted grammatical forms and some relatively unusual word choices, the chosen design created a stringent test of the perceptual salience of **r**-resonances. The point of claims about the communicative salience of phonetic detail of any sort, and subtle, or fine, phonetic detail such as **r**-resonances in particular, is that its distribution in the speech signal must systematically indicate some linguistic structure or communicative function (Hawkins, 2003; Local, 2003). In this first test, we intentionally destroyed this systematicity by presenting in the same session stimuli in which the resonance information in the target word matched that of the rest of the sentence, and an equal number of stimuli in which it did not match. We achieved this by splicing minimal-pair r/l target words into sentences that had contained either the same target word (match) or the other member of the minimal pair (mismatch).

We would expect the following results for native English (L1) listeners. Local English **r**-resonances (those adjacent to or very close to the critical /r/ segment) are strong, so should increase target word intelligibility. If long-domain **r**-resonances are perceptually salient, then they should further increase intelligibility of r-words (match), and decrease intelligibility of the mismatching l-words. In contrast, the lack of long-domain resonances due to an //



suggests that sentences produced with an l-word in the target position would not strongly predict an upcoming //, though they might predict that an /r/ should not follow.

### *3.1. Participants*

Participants were twenty-seven native speakers of English and 27 native speakers of German. Mean ages were 21 years (range 18 - 30) for the English group and 29 years (range 19 - 40) for the Germans. All volunteers reported normal speech and hearing and were students or partners of students at the University of Cambridge. They were reimbursed for their time. Table 2 summarizes age, gender, and the Mill Hill vocabulary score (J. C. Raven, Raven, & Court, 1982) for both groups of participants, together with the German speakers' length of stay in the UK and in other English-speaking countries. German speakers scored slightly lower on the 20-item vocabulary test than English speakers (see Table 2). It should be noted that the native English speakers' mean score of 15.56 was exceptionally high: above the 95<sup>th</sup> percentile in the norms for this age group (J. Raven, Raven, & Court, 1998). The mean score for native German speakers (13.26) was also very high: just below the 90<sup>th</sup> percentile in the norms for native English speakers. The lowest individual German score was at the 75<sup>th</sup> percentile for the age group. It would thus be misguided to interpret the Germans' lower mean score as evidence of a deficient English vocabulary.

INSERT TABLE 2 ABOUT HERE

The native English group all spoke British English, mostly either Standard Southern British English (SSBE) or Received Pronunciation (RP), which is similar to but less widely-spoken than SSBE. Neither accent is rhotic.

The German native speakers spoke a number of regional varieties of standard German and had lived in the UK for an average of 29.25 months (SD: 26.26, range 3-105). All started learning English as a second language at age 11 or later, none having used it regularly outside school lessons. For 19 of the 27, this was their first extended stay in an English-speaking country. Of the remaining eight German participants, one had lived in Australia and seven had already spent at least one year in North America, of whom one, resident in the UK for only 5 months, had attended an English-medium university in Germany. Table 3 shows these details for each individual German participant.

INSERT TABLE 3 ABOUT HERE

### 3.2. Methods

#### 3.2.1. Material

52 experimental sentence-pairs and 63 filler sentences were recorded from one male near-RP speaker who read each sentence five times. The members of each sentence pair differed in only one phoneme, /r/ or /l/, in one word, the target word. For example: *She was sitting there, just watching Terry/telly*. The complete list is in Appendix 2. Each **target word** (*Terry/telly*) contained just the one /r/ or the one /l/. The rest of the utterance (the **sentence base**) contained neither /r/ nor /l/. Thus, each sentence had exactly one /l/ or one /r/ in it, always in the target word. Words with an /l/ (**l-words**) were chosen as contrasts for words with /r/ (**r-words**) because previous work suggests that [l] does not generate long-domain resonances and, in minimal pairs, syllables immediately adjacent to an /l/ are auditorily distinct from those adjacent to an /r/ (Coleman, 2003; Hawkins & Slater, 1994; Heid & Hawkins, 2000; Kelly & Local, 1986). There was no significant difference between the mean frequency of occurrence of the l- and r-target words in the language as measured by the British National Corpus

(Leech, Rayson, & Wilson, 2001) (l-words: mean 10.82, SD 18.52; r-words: mean 13.72, SD 28.66 per 1 million words;  $t[102] = 0.613$ ,  $p = 0.54$ ).

Stimuli were made as follows. A token of a target word was chosen from one of the five tokens of a sentence and spliced into the base of another token of the same sentence to produce a **matched** stimulus: a second *Terry* replaced the original *Terry*. That same target word token was also spliced into the base of the other member of the pair to produce a **mismatched** stimulus: the second *Terry* replaced the original *telly*. Likewise, matched and mismatched stimuli were made for *telly*. This process resulted in four stimuli for each original pair: two matched, and two mismatched, for a total of  $4 \times 52 = 208$  experimental stimuli: 52 matched r-sentences (an r-base with an r-word (rr)) and 52 matched l-sentences (an l-base with an l-word (ll)), and the same number of mismatched sentences (an r-base with an l-word (rl) or an l-base with an r-word (lr)).

With two exceptions noted in Appendix 2, all target words were spliced from the beginning of the first acoustic segment corresponding to the word, to the end of its last acoustic segment, with phonetic contexts chosen to facilitate high quality splicing. Specifically, when the critical word began with an aperiodic sound, it was preceded by a periodic segment, whereas critical words that began with periodic sounds were preceded by aperiodic or silent segments. Examples of splicing: at the boundary between periodicity offset and aperiodicity onset for [s] in cases like *the sprinter/splinter*; in the [p] silence for *the prank/plank*; at the boundary between aperiodicity for [s] and periodicity for l/r in *six rams/lambs*. The same criteria were applied to the ends of words. Tokens were chosen for splicing on grounds of producing good formant continuity together with minimal differences in  $f_0$ , rhythm, rate, and loudness.

Stimuli were checked for technical quality of the splice and naturalness by the third author, who is a trained phonetician and native speaker of English. In most pairs the splice sounded

good, with no strong cues as to the next segment. In three pairs (5, 7, and 8 in Appendix 2) the phone before the splice point carries reasonably strong cues as to whether the next sound is /r/ or /l/: *six rams/lambs*; *biggest rock/lock*; *damaged reef/leaf*. In the case of *six rams/lambs*, the final /s/ indicates place of articulation of the next sound, so a mismatched next sound will almost inevitably be somewhat audible, though it may not be perceptually salient in noise. In the other two cases, normal SSBE coarticulation in the clusters /str, stl, □dr, □dl/ produces differences in the frication/aspiration following the alveolar stop burst. Each of these allophonic differences is, of course, a resonance effect due to the /r/ or the /l/, and so is relevant to the focus of the study. Moreover, the effect is balanced in that if it is audible in one mismatched stimulus, then it is also audible in the other mismatched stimulus of that sentence pair.

Target phonemes were either **word onset** (*rakes*) or **word medial** (*Terry*). Although word onset phonemes occurred either word initially (*rakes, lakes*) or in a consonant cluster (*sprinter, splinter*) both phoneme positions were categorised as onset, and due to a lack of sufficient sample sentences a possible difference between these positions was not further investigated. Thus an important distinction was that, for word-medial target words, the sonorant segment before a word medial /r/ or /l/ always came from the same token as the critical approximant itself. In contrast, for word onset /r/ or /l/, the syllable preceding the target phoneme in the spliced sentences always came from a different utterance, either one which had originally preceded the same r-word or l-word, thus preserving the same resonance quality (matched-splice) or else one which had originally preceded the other member of the target pair (mismatched-splice), thus producing the wrong resonance quality in the syllable immediately adjacent to the critical /r/ or /l/. So in all cases the original acoustic coherence of the target word itself was preserved, but in the word-medial cases, the syllable before the critical /r/ or /l/ was part of the same token, whereas in word-initial cases, it was not. This

distinction is important because resonance cues are normally stronger in the syllable adjacent to the /r/ than in earlier syllables (Heid & Hawkins, 2000).

The recording was made directly onto a Silicon Graphics hard disk using *xwaves*, in a double-walled sound-attenuated chamber with a Sennheiser MKH 40 P 48 microphone about 30 cm from the speaker's mouth. Sampling frequency was 22.05 kHz. After removing dc offset, the sound files were converted to wav files. Splicing was done using Praat.

After splicing, the sentences were masked with British 12-talker babble. The babble was made by mixing independent recordings of twelve native British English speakers of various accents reading a book or newspaper of their choice in the Cambridge University Phonetics Laboratory sound-treated booth. Sampling rate was 22.05 kHz. All single-voice recordings of the babble had equal long-term RMS amplitude values to ensure that no voice stood out.

Target sentences and noise were combined at +1 dB SNR for English speakers and +4 dB for German speakers. Pilot tests indicated that these different SNRs should equate overall intelligibility for the two groups of listeners. This was confirmed by an independent-group t-test on the overall intelligibility of target words ( $t[52] = 1.74$ ,  $p > 0.05$ ). Native English speakers achieved an intelligibility of 61.4% (s.d. = 0.10), native German speakers of 55.7% (s.d. = 0.14). The babble began 2 s before sentence onset to allow listeners to adapt to the background noise; this improves speech intelligibility and memory for speech (Heinrich et al., 2008; Wagener & Brand, 2005). The same noise segment was used for all four versions (ll, lr, rl, rr) of a sentence pair, with different noise segments used for different sentences.

### 3.2.2. Procedure

After volunteers gave written consent, they completed the Mill Hill Vocabulary Test (J. C. Raven et al., 1982) and underwent audiometric testing. The audiogram measured hearing

sensitivity (left ear then right) at 6 frequencies (0.25, 0.5, 1, 2, 4, 8 kHz) in accordance with ANSI (1996) standards. A GSI 61 clinical audiometer was used. Mean thresholds for all frequencies, ears and groups were < 10 dB HL.

Participants were tested individually in a double-walled IAC booth in Cambridge University Phonetics Laboratory. They heard the stimuli through Sennheiser HD 250 headphones from a PC sound card via a TEAC A-X1000 amplifier with fixed-level output. All completed a 10-sentence practice session prior to the test session. The practice sentences were selected from the unspliced filler sentences and were the same for all participants.

In the experiment proper, stimuli were counterbalanced across 4 groups such that each listener heard each of the 52 test sentences in only one condition, that is as a matched or mismatched l- or r-sentence. In each language group, sentence orders 1-3 were each listened to by seven participants, while the remaining six participants listened to sentence order 4. All four orders were randomly generated with the only constraint that all four types of sentences (ll, lr, rl, rr) were distributed roughly equally throughout the list. Sentences were between 6 and 15 words long (mean: 10.3, s.d.: 2.2). For each trial, a mask on the computer screen displayed the first 1-5 words of the sentence (mean: 2.3, s.d.: 1.2) while the sound file was played. Participants listened to each sentence and then typed the remainder of the sentence into the mask. The number of words in the mask was adjusted such that, on average, the target word occurred in the 6th position of the words the participants typed (mean: 6.1, s.d.: 1.6). The task was self-paced and took between 20 and 30 minutes to complete. German speakers completed the task with the same English sentences.

### 3.3. Results

Total Ns were adjusted to account for the few (0.2%) missing data points. All results are presented as percent correct target words. There were no differences in overall intelligibility between word-onset and word-medial target words for either native English or the German listeners. Patterns of intelligibility within the onset and medial sets differed, however, so were analysed separately. In all analyses, percent correct identified words for each of the four splicing conditions (ll, rr, lr, rl) is shown separately according to whether the target phoneme was (a) in the word onset or (b) word medial. Unless otherwise stated, the main statistical analyses were 2-base (l/r) x 2-match (match/mismatch) fixed-factor repeated measures ANOVAs; posthoc tests were SIDAK-corrected.

#### 3.3.1. The role of resonance information

Figure 2A shows the percentage of target words correct for English native speakers. When the critical /l/ or /r/ occurred at word onset, there was a marginally significant interaction between l/r base x match/mismatch ( $F[1, 26] = 3.92, p = 0.06, \text{partial } \eta^2 = 0.13$ ) but no main effects of either sentence base (l/r) or target word match/mismatch (both  $F[1, 26] < 1$ ). This interaction occurred because on the whole, l-words were marginally more intelligible than r-words in word onset conditions.

INSERT FIGURE 2 ABOUT HERE

When the critical /l/ or /r/ was word medial, there again was an interaction between base and match, this time clearly significant ( $F[1, 26] = 18.64, p < 0.001, \text{partial } \eta^2 = 0.42$ ). Moreover, as before, neither main effect was significant (l/r-base:  $F[1, 23] = 2.08, p = 0.16$ ; match/mismatch:  $F[1, 26] = 1.45, p = 0.24$ ). However, here, the nature of the interaction is very different from the word-onset case; it arises because l- and r-base lead to different

patterns of intelligibility: in an l-base, it is unimportant whether the target word represents a match (l-word) or mismatch (r-word) to the base; in both cases the words are equally intelligible ( $t[26] = 1.1$ ,  $p = 0.30$ ). For an r-base, on the other hand, the match to the following target word plays an important role so that the matched rr-condition was significantly more intelligible than the mismatched rl ( $t[26] = 3.4$ ,  $p = 0.002$ ). When results were averaged over sentence-base conditions, word-medial r-words were more intelligible than l-words ( $rr + lr > ll + rl$ ,  $t[26] = 4.32$ ,  $p < 0.001$ ). Further, within the lr condition alone, word-medial target words were more intelligible than word-onset words ( $t[26] = 3.00$ ,  $p = 0.006$ ).

Given that word frequencies did not differ between l- and r-words in either word-onset or word-medial positions (both t-values  $< 1$ ), it is unlikely that differences in word familiarity caused this position-dependent difference between l- and r-words. It is more likely that the acoustic cues provided by the r-base can help explain the different patterns of intelligibility between l- and r-words in matched and mismatched splicing conditions. Hence, we compared the effects of splice point and sentence-base separately for l- and r-bases by running splice-point (word-onset, word-medial) by match (match/mismatch between base and target word) repeated-measures ANOVAs separately for l-base conditions (word-onset: ll, lr; word-medial: ll, lr) and r-base conditions (word-onset: rr, rl; word-medial: rr, rl).

In the l-base case, the only significant effect was that of splice-point ( $F[1, 26] = 4.39$ ,  $p = 0.05$ ) indicating that on the whole, words were more intelligible when there was an acoustically undisturbed syllable between the splice-point and the critical phoneme (word-medial conditions). Neither the main effect of match ( $F[1, 26] < 1$ ,  $ll = lr$ ) nor the interaction between splice-point and match ( $F[1, 26] = 2.32$ ,  $p = 0.14$ ) were significant, even though Figure 2A might suggest otherwise.



The pattern looked very different for r-bases: the analysis showed no main effect of splice-point ( $F[1, 26] = 1.61, p = 0.22$ ) or match ( $F[1, 26] < 1$ ), but there was a significant interaction ( $F[1, 26] = 4.37, p = 0.047$ ). The interaction indicated that matching an r-base with an r-target word had no effect on intelligibility when the critical phoneme was in the word onset (onset:  $rr = rl, t[26] < 1$ ) but matching did increase intelligibility of r-words over l-words when the critical phoneme was word-medial ( $rr > rl, t[26] = 3.38, p = 0.002$ ). It seems that the undisturbed syllable prior to the critical phoneme is of crucial importance for the intelligibility of r-words.

This difference in r-word intelligibility between word-medial and word-onset conditions is probably due to differences in the structure of the stimuli. In order to preserve the original acoustic coherence of all target words, the word-onset and word-medial conditions differ in the type of acoustic-phonetic information immediately preceding the critical /r/ or /l/ segment. In word-medial target words, the syllable before the critical segment comes from the same token as the critical segment itself, rather than being spliced from another token of the same utterance type, and the segment adjacent to the critical /r/ or /l/ segment is normally a vowel (see Appendix 2). In word-onset target words, in contrast, the preceding vowel always comes from a different token of the utterance, and there are one or more obstruent consonants between that vowel and the critical segment in the target word. Compared with word-onset targets, then, word-medial targets could be expected to bear stronger local coarticulatory influences: maximally coherent cues to r-resonances in the matched rr cases, and only somewhat incoherent cues in the mismatched lr cases. Word-onset target words would be expected to have somewhat weaker acoustic-perceptual cues to /r/ immediately before the critical /r/ segment, because the critical /r/ and the vowel in the preceding syllable are separated by one or more consonants and the splice point. It is presumably this difference which was responsible for the 14% increase in intelligibility of word-medial over word-onset

r-words, from 55.0% to 69.4 % (word-medial (rr+lr) > word-onset (rr+lr),  $t[26] = 4.72$ ,  $p < 0.001$ ).

Figure 2B shows the equivalent data for German speakers. For targets with the critical phoneme in the word onset, there was a significant interaction between l/r base x match/mismatch ( $F[1, 26] = 13.1$ ,  $p = 0.001$ , partial  $\eta^2 = 0.34$ ) but no significant main effects (both  $F$ 's < 1). The interaction was mostly driven by the difference between the two r-base conditions, rr and rl, as would be expected, but the difference was in the unexpected direction: mismatched target words were more intelligible than matched ones (rl > rr,  $t[26] = 2.15$ ,  $p = 0.04$ ). The difference between the two l-base conditions was in the expected direction, with matched l-words more intelligible than mismatched r-words (ll > lr) and approached significance in posthoc tests ( $t[26] = 1.83$ ,  $p = 0.08$ ).

In complete contrast to the English group, the German listeners showed no significant differences in intelligibility amongst targets with word-medial phoneme locations. (Main effect of match, and the base x match interaction,  $F$ s < 1 ; base,  $F[1, 26] = 2.80$ ,  $p = 0.11$ , partial  $\eta^2 = 0.1$ ).

In both word-onset and word-medial conditions, the pattern of results for German listeners was thus quite different from that for English listeners. The unexpected patterns encouraged closer examination of the German results. Recall that the German listeners differed greatly in the time they had spent in the UK before testing (3-105 months, Table 3). A median split of the group based on duration of stay in the UK gave a value of 20 months, with 14 listeners having spent 20 or fewer months in the UK (the short-stay group), while the other 13 had spent more than 20 months in the UK (the long-stay group). The mean duration of stay in the UK for the short- and long-stay groups was 10 months (s.d.: 4.19) and 50 months (s.d.: 24.02)

respectively. Figures 2C and D show percent correct target words for the short-stay and long-stay groups separately.

The long-stay group showed no sensitivity to **r**-resonances: there were no significant differences amongst any of the conditions in either word-onset or word-medial phoneme location analyses, although the interaction between base and match in the word-onset analysis did approach significance, mirroring the  $rl > rr$  pattern found for the entire German group (Figure 2B). (Word-onset phoneme location: main effects,  $F$ 's  $< 1$ ; base  $\times$  match:  $F[1,12] = 4.30$ ,  $p = 0.06$ , partial  $\eta^2 = 0.26$ . Word medial phoneme location: main effects,  $F$ 's  $< 1$ ; base  $\times$  match  $F[1, 12] = 2.63$ ,  $p = 0.13$ , partial  $\eta^2 = 0.18$ ).

The response pattern of the short-stay group, with 20 or fewer months continuous recent exposure to British English, resembled that of the long-stay German listeners in some ways, but that of the native English listeners in other ways. Like both these other groups, the short-stay German listeners produced no significant differences in target word intelligibility dependent on either  $l/r$  base or match/mismatch alone. (In both word-onset and word-medial phoneme location ANOVAs, all main effect  $F$ s  $< 1$ .)

However, the short-stay group's base  $\times$  match interaction was significant in both onset and medial analyses. For word-onset stimuli, the interaction was like that for the long-stay Germans, i.e. caused by an unpredicted 15% advantage for  $l$ -words over  $r$ -words when in an  $r$ -base:  $rl > rr$  ( $F[1, 13] = 8.872$ ,  $p = 0.011$ , partial  $\eta^2 = 0.41$ ). In other words, the entire German group showed this  $rl > rr$  pattern in the word-onset condition, regardless of length of stay in the UK. The opposite direction of difference for the short-stay group's onset  $ll > lr$  is almost the same absolute size (14%) but is not significant ( $t[13] = 1.59$ ,  $p = 0.14$ ). Although this difference is consistent for both long- and short-stay German listeners, it only approaches

significance for the entire German group, as noted above (Figure 2B). More work would be needed to examine this further.

But the opposite pattern held for the short-stay group's word-medial stimuli: target words were more intelligible in the matched rr condition than in the mismatched rl condition (base x match interaction:  $F[1, 13] = 4.61$ ,  $p = 0.05$ , partial  $\eta^2 = 0.26$ ), with the  $rr > rl$  difference approaching significance in a posthoc t test ( $t[13] = 1.96$ ,  $p = 0.07$ ). Thus, although the direction of difference was contrary to the long-stay German listeners' pattern, it was the same as that of the native English listeners, although the absolute difference is much greater for the native English group: a 24% drop in intelligibility between rr and rl target words for the English, compared with only a 10% difference for the short-stay Germans.

Most interestingly though, and in contrast with the English speakers, short-stay German speakers did not take advantage of the strong r-resonance cues in the syllable immediately before the /r/ segment. Rather, all intelligibility benefit stemmed from the longer-lasting r-resonance information in the sentence base ( $rr > rl$ , see above), while local r-resonance cues did not appear to be important (word-medial  $ll = lr$  ( $t[13] = 0.41$ ,  $p = 0.69$ )). This point is taken up further in the Discussion.

The absence of significant effects in the long-stay German group could not be explained by a lack of power due to the relatively small group size for two reasons. First, the short-stay German group contained the same number of participants, yet showed a significant interaction effect. This suggests that group size was sufficient. Second, we calculated the number of data points needed for an effect of the size shown by the long-stay German group to reliably reach significance, using the freeware G\*Power (Faul, Erdfelder, Lang, & Buchner, 2007). We first computed the effect size of the interaction based on Cohen's d (Cohen, 1988),  $d = 0.30$ . The number of data points needed to detect this effect size reliably, that is with a power of 0.8, is

N=91. This number is much higher than any of the sample sizes tested in the current experiment. Hence, we can be confident that with the sample sizes employed in the current experiment, an effect of the size seen in the long-stay German group would have been very unlikely to reach significance.

INSERT TABLE 4 ABOUT HERE

If long-stay German listeners were not using **r**-resonances, were they using anything else? An obvious possibility is word frequency. Table 4 shows Pearson product-moment correlations between each target word's frequency according to the British National Corpus (BNC; Leech et al., 2001) and its intelligibility score achieved by the three groups of listeners: English, and short- and long-stay Germans. Whereas there was only one significant correlation between target word identification and word frequency for the English listeners (l1), and the short-stay German listeners (r1), all four conditions correlated significantly with word frequency for the long-stay German listeners. It seems that, in this task, German speakers with about 2-8 years' immersion in British English use word frequency more than the other two listener groups to help them identify words in a sentence. Conversely, the other two listener groups used the **r**-resonances in the physical signal more, despite differing in native language.

### 3.3.2. *Effects of splicing*

The intelligibility advantage due to **r**-resonances presumably stems from some sort of perceptual coherence, probably language-specific. Splicing itself can be expected to introduce some incoherence. As noted in Section 3.2.1, these stimuli were spliced manually with great care to produce good formant continuity together with minimal differences in  $f_0$ , rhythm, rate, and loudness. Nevertheless the use of natural speech meant that small discontinuities were inevitable and could be expected to reduce intelligibility.

To estimate the degree of incoherence introduced by splicing, we scored word intelligibility for three words preceding the target. Those words included the word immediately preceding the splice point, and two words earlier in the sentence. As far as possible, the words were chosen to be independent of each other, and to be relatively stressed in the sentence. They are the underlined words in Appendix 2. These data are shown in Figure 3, with word-onset and word-medial target word conditions in the left and right panels respectively. Each target-word condition was subjected to a 4 x 2 mixed-measures ANOVA with word location and language group as within- and between-group factors.

In both word-onset and word-medial, conditions, and for both language groups, the three words heard before the splice point (-1, -2 and -3 in Figure 3) were much more intelligible than the target word (+1) which of course came after the splice point. The main picture, then, is clear: despite the care taken, splicing these natural tokens did affect intelligibility for both native and non-native listeners. (Main effect of word in the onset analysis:  $F[3, 156] = 84.14$ ,  $p < 0.0001$ , partial  $\eta^2 = 0.62$ ; in the medial analysis:  $F[3, 156] = 242.71$ ,  $p < 0.0001$ , partial  $\eta^2 = 0.82$ .) Sidak-corrected posthoc t-tests confirmed that the spliced target word (+1) was considerably less intelligible than any of the other words, for each language group. It is worth mentioning that the majority of target words (34 pairs) were not sentence-final and thus had two splice points, one at the beginning and one at the end of the word. These two sources of acoustic discontinuity could be expected to have quite a significant combined effect.

INSERT FIGURE 3 ABOUT HERE

Figure 3 also shows that our effort to equate native English and German listeners with regards to overall intelligibility of target words was reasonably successful. In the word onset condition, English listeners outperformed German listeners before the splice point only by about 6% (English: 85%; German: 79%), while after the splice point performance was

virtually identical for both groups (English: 59%, German: 58%). The 6% intelligibility difference in the words before the splice point however was large enough to be statistically significant. ( $F[1, 52] = 7.56$ ,  $p = 0.008$ , partial  $\eta^2 = 0.13$ ). In the word medial condition intelligibility before the slice point was well matched for both groups and did not lead to significant group differences ( $F[1, 52] = 2.71$ ,  $p = 0.11$ , partial  $\eta^2 = 0.05$ ).

The most interesting part of the 4 word x 2 language group analysis is the significant interaction between r/l-Word and Language Group that was present in the word-medial condition (medial:  $F[3, 156] = 3.75$ ,  $p = 0.012$ , partial  $\eta^2 = 0.07$ ; onset:  $F < 1$ ). This interaction was due to the fact that native English listeners correctly perceived more word-medial target words than did non-native listeners. Figure 2 shows that the listening advantage for native listeners is due mainly to better intelligibility for lr and rr-words, a result that, as expected, underscores yet again how helpful local **r**-resonances are to native listeners.

We conclude, then, that splicing adversely affected word intelligibility for both listener groups. Interestingly, though, while native and non-native listeners did not differ in their effect of splicing for word onset conditions, non-native listeners suffered considerably more in word-medial conditions presumably because they did not use the cues provided by resonances to the extent that native listeners did.

#### **4. Discussion**

The present experiments demonstrate for the first time that **r**-resonances in the acoustic signal can affect word intelligibility in unrestricted natural read speech heard in adverse listening conditions with a free response task. Previous experiments demonstrated perceptual salience of **r**-resonances only for synthetic speech (Hawkins & Slater, 1994; Tunley, 1999), or with natural speech but using a single carrier phrase and a two-alternative forced choice task

(West, 1999b). The present conditions introduced a far greater degree of natural variation and hence unpredictability.

Under these stringent test conditions, native SSBE speakers tended to use local **r**-resonance information in word medial but not word onset splicing conditions. As explained in Section 3.3.1, the fact that **r**-words were more intelligible when they were word-medial (word-medial  $l_r >$  word-onset  $l_r$ ), indicates that an undisrupted sonorant segment immediately adjacent to the critical **/r/** segment facilitates **r**-word identification, as would be expected given the strong **r**-colouring measured in the nuclei of syllables immediately adjacent to an **/r/** segment (Heid & Hawkins, 2000; Tunley, 1999). Further work is needed to assess whether this large local effect is due mainly to the nature of the phonetic segments abutting the **/r/** segment, or whether other factors are influential. These other influences might include statistical properties of lexical structure and sequential predictability, and, importantly, the acoustic integrity of the natural signal—its acoustic-perceptual coherence—which splicing disrupts.

When the target phoneme segment was located immediately after the splice point (word onset conditions) the presence or absence of **r**-resonances in the sentence base did not affect intelligibility. Given that local **r**-resonances affected intelligibility in this task by 10-20% (Figure 2), and that there was a 20-30% decrease in intelligibility for both native and non-native listener groups for the word after the splice point (Figure 3), it seems reasonable to assume that the acoustic disturbance introduced by splicing (at one or both ends of the word) may have masked any long-domain resonance effect. Further exploration of these observations requires a design that does not disrupt the signal, and/or a more sensitive measure of responses to **r**-resonances.

There is no unambiguous evidence in these data for the perceptual salience of long-domain **r**-resonances in this task. To ideally demonstrate their salience, **l**-bases should not be especially



predictive of l-words, whereas r-bases should facilitate intelligibility of r-words, and decrease intelligibility of l-words by virtue of misleading the listener in the preceding base. There is evidence for this latter misleading effect of a preceding r-base, in the English listeners' large and strongly significant difference in intelligibility between matched r-words and mismatched l-words ( $rr > rl$ ) word-medially. That the effect is largely interference rather than facilitation follows from the lack of significant difference between the other three word-medial conditions (Figure 2A). The significant positive correlation between  $ll$  and word frequency for English listeners (Table 4) further suggests that, as expected, these listeners are not using the l-base to predict an upcoming word because there are no long-domain resonances specific to English [l]. Suggestive evidence for a facilitatory effect of an r-base come from the pattern of word-medial data for German short-stay listeners (Figure 2C): intelligibility is best only when the short and long-domain r-resonances are coherent ( $rr > lr = rl$ ). However, although the interaction is significant, none of the individual comparisons achieve statistical significance. Thus, nothing stronger can be said about the perceptual salience of long-domain resonances from the present data. Yet, this issue is worth further investigation because the same patterns reliably recur in experiments, not reported here, which use the same stimuli in the same stringent design, but different listening conditions.

For native German listeners, the use of r-resonance information seemed to be related to the amount of time spent in the UK. The group of native German listeners who had spent less time in the UK seemed to use English r-resonances to aid word identification, while the group with the longer exposure to SSBE did not. Moreover, quite unlike native English listeners, the non-natives only use long-domain cues, ignoring the properties of the vowel immediately adjacent to the /r/ segment which are probably the primary influence for native English listeners. That is, for word-medial targets, short-stay German listeners showed the same  $rr > rl$  advantage as native English listeners (albeit smaller in absolute terms) but no advantage of  $lr$  over  $ll$ . And short-stay Germans had no advantage for medial over onset  $lr$ . Thus they were

not using the local resonance information. Long-stay non-natives, in contrast, had no significant differences for these comparisons, indicating that they took no advantage of either long-domain or local **r**-resonances. Instead, they used lexical frequency more than both the native listeners and the short-stay Germans (Table 4, correlation coefficients).

Higher correlation coefficients for long-stay Germans compared with short-stay Germans presumably reflect greater facility in processing spoken English, although in this case it does not simply reflect better vocabulary, for the two groups had very similar (and very good) English vocabulary scores: means 13.07 (s.d. 2.37) and 13.46 (s.d. 3.23) for short- and long-stay groups respectively.

However, these higher correlations cannot mean that the long-stay Germans had greater facility in English than the native English listeners. It is more likely that long-stay Germans learn to ignore English **r**-resonance cues altogether. The reason presumably lies in what the Germans bring from L1 to the L2 listening task in Experiment 2, in which acoustic cues to /r/ and /l/ had been experimentally manipulated so that they were unsystematic. Recall that the main acoustic difference between German and English /r/ lies in the frequencies of F3 and F4 relative to surrounding segments; F3 and F4 are high for German and low for English. In both languages, F1 frequency is slightly higher for /r/ than for comparable phonological categories, and F2 is always low for German and sometimes low for English. While high F1 is not a characteristic property of [l], a high F3 is. So whereas German [ʀ] and [l] have similarly high F3 frequencies, English [ɹ] and [l] contrast, with F3 low in [ɹ] but high in [l]. As [l] and [ʀ] do not contrast in F3 frequency in German, German native speakers presumably focus little attention on F3 when distinguishing these two consonants. Consequently, they may also not use F3 frequency when listening to English [l] and [ɹ].

The literature on English **r**-resonance reviewed in the Introduction, and the data of Experiment 1 and of Lodge (2003) confirm that, in both languages, formant differences are most strongly represented in vowels close to the conditioning /r/ or /l/ segment. Note that despite the prominence of the contrast in acoustic detail between /l/ and /r/ in English and adjacent vowels, German listeners do not make use of local **r**-resonances.

Why then would short-stay listeners seem to use the weaker long-domain resonances? One possibility is that family resemblance between the two sounds may play a role (e.g., Lindau, 1985). If the Germans recognize general but weak acoustic properties of a rhotic articulation, they may use them in adverse listening conditions without noting the difference in the fine detail of the pattern of formant frequencies between English and German. F2 frequency is the most reliably lowered in German [ʀ], while effects on F2 frequency in English [ɹ] tend to be small and may be as likely to be higher as lower (West, 1999b). In contrast, F3 is strongly and reliably affected in English [ɹ] (e.g., Heid & Hawkins, 2000, West, 1999a), but not in German (this paper). As these L2 listeners become more familiar with spoken English, they presumably notice that although the general effect is /r/-like, the phonetic detail is “wrong”, especially with respect to F2 versus F3 and higher formants. When this new knowledge is systematised with respect to the conditioning /r/, it causes confusion. This type of conflict between partially-overlapping categories in L1 and L2 is thought to present a particularly difficult situation for L2 learners, cf. the Perceptual Assimilation Model (Best, 1994; Best, Roberts, & Goodell, 2001).

In summary, neither short- nor long-stay German listeners in this study seem to have learned to use short-domain English cues, and the long-stay Germans seemed to have learned to ignore the long-domain cues too. We interpret this as due to the presence of both congruent and conflicting acoustic cues to rhoticity in German and English.

Presumably, with extended exposure, at least some non-native listeners learn to use English long-domain **r**-resonances again, perhaps in conjunction with the stronger short-domain local cues with which they are compatible. In a word-segmentation task, Ito and Strange (2009) found that native Japanese speakers living in the USA took 9-10 years to learn to discriminate harder English phonological contrasts. Only four of our Germans had lived in an English-speaking country for anything approaching 9 years (108 months), and for three of them this included at least a year in North America, where velarization of /l/ as well as /r/ is the norm, and may affect listeners' predictive use of **r**-resonances. Whether Germans living in the UK start to use **r**-resonances again therefore remains an open question.

An issue that can only be speculated about at this point is whether long-domain **r**-resonances are found in all varieties of English that use [□], and whether all long-domain **r**-resonances are perceptually salient. For example, a rhotic accent might exhibit long-domain **r**-resonances, but they might be unimportant perceptually because /r/ is so prevalent in English that its long-domain manifestations would be found throughout most utterances in a rhotic accent. Equally, long-domain **r**-resonances in nonrhotic accents might vary in perceptual salience dependent upon how velarised the /r/ is. The degree of velarisation of syllable-onset /r/ depends partly on its relationship with onset /l/ in the phonological system, as briefly discussed by Hawkins and Nguyen (2004), and in detail by Carter (1999, 2002, 2003), Kelly and Local (1986; 1989) and Coleman (2003).

Presumably, in nonrhotic accents with dark (i.e. velarised) /r/ and clear (non-velarised) /l/, the velarisation is perceptually salient because it contrasts with the rest of the system, which is less velarised. Thus we might predict that **r**-resonances would be less perceptually salient in rhotic varieties of American English, where both onset /r/ and /l/ tend to be strongly velarised

(or the auditory equivalent, Secord et al., 2007), than in English English varieties like SSBE, where onset /l/ is not velarised at all (Cruttenden, 2001 and references cited earlier).

This speculation is supported in our present data by the fact that the crucial word-medial rr-rl intelligibility difference was comparable for the native English listeners and for the 9 short-stay German listeners who had not lived in the USA (15.7% for native English, 16.1% for Germans). Whereas the 5 short-stay Germans who had also lived in the USA showed almost no difference (1.7%) between rr and rl. Conversely, differences between word-onset rr and rl were large and in the opposite direction (26.6%) for those who had lived in the USA, and small (5.2%) for those who had not. These five German listeners who had lived in the USA were not removed from the main analysis because the loss of statistical power from the reduced N meant that the overall difference failed to achieve significance. To include these L2 learners who had also lived in the US seems acceptable, since their results work against our hypothesis. However, we can conclude that even non-recent exposure to US English, or perhaps rhotic English, can affect L2 listeners' processing of phonetic detail due to r-resonances. It would be worthwhile in future work to explore the influence of rhoticity and widespread velarisation on the production and perception of long-domain resonance in the English liquid system.

These data provide evidence for individual differences in listening styles due to experience. Understanding speech requires the listener to match the incoming physical signal to expected patterns that make linguistic sense, be those patterns memories of episodes, abstract categories of various types, or both (e.g., Hawkins, 2010; Moore, 2007). Whereas all listening requires such dual use of prior knowledge and physical signal, the extent to which prior knowledge is used presumably depends heavily on signal quality, listener characteristics and the listening task. For example, Luce and colleagues (Luce, Feustel, & Pisoni, 1983; Luce & Pisoni, 1998) and Sommers (1996) demonstrated how these factors interact in the case of

single word perception: When the speech input was compromised, either because of poor signal quality (background noise), listener constraints (age-related hearing loss) or both, intelligibility scores decreased most for those words that had many phonetically similar competitors. They argued that this result occurred because the imperfect sound pattern of the input signal partially activated many similar and highly confusable words in the mental lexicon from which the listener subsequently had to choose. For words with few phonetic competitors the quality of the physical signal was less important because there was less potential for confusion. Extrapolating from the intelligibility of single words to that of whole sentences, one can argue that in addition to lexical variables, syntax and context play a crucial role in the activation and choice of lexical alternatives for a target word.

Secondly, both Sommers (1996) and Luce et al. (1983) argued that monitoring, updating and choosing between the activated lexical alternatives consumes considerable cognitive resources. If we assume that cognitive resources are limited (Kahneman, 1973), and are crucial for processes such as understanding, integrating and memorizing heard information (e.g., Craik, Govoni, Naveh-Benjamin, & Anderson, 1996), then a sparse use of these resources for the actual signal processing would be efficient. When the task is to write down spoken words, a clearly-spoken signal with no unpredictable or ambiguous words should demand fewer cognitive resources than a less clearly-spoken signal heard in noise. In this latter case, word predictability, mediated by knowledge of the language as a whole, will provide a good fit between the physical signal and the linguistically probable item. When a word is relatively unpredictable and is heard in noise (as were the key words in these stimuli), then its activation level will be relatively indistinguishable from those of many other lexical items. In this case, additional cognitive resources are presumably demanded to choose a match that fits both the physical signal and the expected linguistic content.

There are many ways in which limited knowledge of the language might put L2 listeners at a disadvantage in this process: (1) they possess less linguistic knowledge of L2, which in addition might be (2) less readily available or take more effort to activate. (3) Their smaller and probably less 'retrievable' L2 lexicon might provide them with fewer and less appropriate lexical alternatives which in turn (4) might make it more difficult for the listener to choose an acceptable lexical alternative.

Another way of expressing this same point, but taking a more agnostic theoretical position on what is cognitively costly, what is knowledge, and what is top down, is that the more the physical signal contributes to a lexical decision, the higher the probability that the decision will be right. The present data suggest that the long-domain and short-domain coherence of the signal both influence perceptual decisions, especially for non-native listeners. It would be interesting to see if non-native listeners reach a stage when they possess the 'higher-level' linguistic knowledge, but ignore it in favour of cues in the physical signal (which may also have been restructured in a new knowledge system about the L2 language) or whether intelligibility always remains more cognitively driven in non-native speakers.

**ACKNOWLEDGEMENTS**

Funded by an ERA-AGE (FLARE) grant (RG49525) to the first author, and a Marie Curie Research Training Network grant, Sound to Sense (S2S: MRTN-CT-2006-035561- S2S) to the third. Experiment 1 was conducted by the second author as part of a Master's thesis at the Universität des Saarlandes. We thank Pia Rübiger for helping with stimulus preparation and data collection, and Bill Barry for discussion. Data for the English listeners comprised part of a paper presented at the 2009 International Conference on Speech and Language Processing, Brighton, UK.



**REFERENCES**

- ANSI. (1996). *American National Standard Specification for Audiometers S3.6-1996*.
- Barry, W. J. (1995). Schwa vs schwa plus /r/ in German  
*Phonetica*, 52(3), 228-235.
- Bell-Berti, F., & Harris, K. S. (1982). Temporal patterns of coarticulation - lip rounding.  
*Journal of the Acoustical Society of America*, 71(2), 449-454.
- Benguerel, A. P., & Cowan, H. A. (1974). Coarticulation of upper lip protrusion in French.  
*Phonetica*, 30(1), 41-55.
- Best, C. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The Development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA/London: MIT Press.
- Best, C., Roberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system.  
*Journal of the Acoustical Society of America*, 109(2), 775-794.
- Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 121(4), 2339-2349.
- Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica*, 86(1), 117-128.
- Carter, P. (1999). Abstractness in phonology and extrinsic phonetic interpretation: The case of liquids in English. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (Eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences* (Vol. 1, pp. 105-108). Berkeley, CA: University of California.
- Carter, P. (2002). *Structured variation in British English liquids: The role of resonance*. Ph.D. dissertation, University of York.

- Carter, P. (2003). Extrinsic phonetic interpretation: Spectral variation in English liquids. In J. Local, R. Ogden & R. Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. (pp. 237-252). Cambridge, UK: Cambridge University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: Lawrence Erlbaum.
- Coleman, J. (2003). Discovering the acoustic correlates of phonological contrasts. *Journal of Phonetics*, 31, 351-372.
- Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology-General*, 125(2), 159-180.
- Cruttenden, A. (2001). *Gimson's Pronunciation of English* (6th ed.). London: Arnold.
- Cutler, A., Lecumberri, M. L. G., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *Journal of the Acoustical Society of America*, 124(2), 1264-1268.
- Darwin, C. J., & Gardner, R. B. (1985). Which harmonics contribute to the estimation of first formant frequency? *Speech Communication*, 4(1-3), 231-235.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Grunke, M. E., & Pisoni, D. B. (1982). Some experiments on perceptual-learning of mirror-image acoustic patterns. *Perception & Psychophysics*, 31(3), 210-218.
- Hawkins, S. (1995). *Arguments for a nonsegmental view of speech perception*. Paper presented at the XIIIth International congress of phonetic sciences, Stockholm.
- Hawkins, S. (1996). Perceptual modeling of connected speech. In A. P. Simpson & M. Pätzold (Eds.), *Sound patterns of connected speech: Description, models and explanation*, *Arbeitsberichte nr 31* (pp. 173-180). Kiel: Universität Kiel: Institut für Phonetik und digitale Sprachverarbeitung.

- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31(3-4), 373-405.
- Hawkins, S. (2010). Phonological features, auditory objects, and illusions. *Journal of Phonetics*, 38(1), 60-89.
- Hawkins, S., & Nguyen, N. (2004). Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *Journal of Phonetics*, 32(2), 199-231.
- Hawkins, S., & Slater, A. (1994, 18-22 September). *Spread of CV and V-to-V coarticulation in British English: implications for the intelligibility of synthetic speech*, Yokohama, Japan.
- Heid, S., & Hawkins, S. (2000). An acoustical study of long-domain /r/ and /l/ coarticulation, *Proceedings of the 5th seminar on speech production: models and data*. (pp. 77-80.). Kloster Seeon, Germany.
- Heinrich, A., Schneider, B. A., & Craik, F. I. M. (2008). Investigating the influence of continuous babble on auditory short-term memory performance. *Quarterly Journal of Experimental Psychology*, 61(5), 735-751.
- Huggins, A. W. F. (1972a). Just noticeable differences for segment duration in natural speech. *Journal of the Acoustical Society of America*, 51(1270-1278).
- Huggins, A. W. F. (1972b). On the perception of temporal phenomena in speech. *Journal of the Acoustical Society of America*, 51, 1279-1290.
- 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(4), 2348-2360.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kelly, J., & Local, J. K. (1986). Long-domain resonance patterns in English. *Proceedings of IEE International Conference on Speech Input/Output: Techniques and Applications*. London: Institution of Electrical Engineers. Conference Publication 258: 304-309.
- Kelly, J., & Local, J. K. (1989). *Doing Phonology*. Manchester: Manchester University Press.
- Ladefoged, P. (2001). *A Course in Phonetics*. Fort Worth, Texas: Harcourt College Publishers

- Laver, J. (1994). *Principles of Phonetics*. New York: Cambridge University Press.
- Lecumberri, M. L. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *Journal of the Acoustical Society of America*, *119*(4), 2445-2454.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. Harlow: Longman.
- Lindau, M. (1985). The story of /r/. In V. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. (pp. 157-168). Orlando: Academic Press.
- Local, J. (2003). Variable domains and variable relevance: interpreting phonetic exponents. *Journal of Phonetics*, *31*(3-4), 321-339.
- Lodge, K. (2003). A declarative treatment of the phonetics and phonology of German rhyml /r/. *Lingua*, *113*(10), 931-951.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence-intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476-490.
- Luce, P. A., Feustel, T. C., & Pisoni, D. B. (1983). Capacity demands in short-term-memory for synthetic and natural speech. *Human Factors*, *25*(1), 17-32.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1-36.
- Machelett, K. (1996). Zur Lautunterscheidung innerhalb der Lautklassen: Bestimmung der Artikulationsstelle. In *Das Lesen von Sonagrammen V1.0 - Kapitel III*. Universität München.: Institut für Phonetik und Sprachliche Kommunikation.
- <http://www.phonetik.uni-muenchen.de/studium/skripten/SGL/SGLKap3.html#/r/>.
- Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech Language and Hearing Research*, *40*(3), 686-693.

- Moore, R. K. (2007). Spoken language processing: Piecing together the puzzle. *Speech Communication, 49*(5), 418-435.
- Nieto-Castanon, A., Guenther, F. H., Perkell, J. S., & Curtin, H. (2005). A modeling investigation of articulatory variability and acoustic stability during American English /r/ production. *Journal of the Acoustical Society of America, 117*(5), 3196-3212.
- Ogden, R. (2009). *An Introduction to English Phonetics*. Edinburgh: Edinburgh University Press.
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven Manual: Section 5 Mill Hill Vocabulary Scale*. Oxford: Psychologists Press.
- Raven, J. C., Raven, J., & Court, J. H. (1982). *Mill Hill Vocabulary Scale*. Oxford: Oxford Psychologists Press.
- Remez, R. E. (1994). A guide to research on the perception of speech. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 145-172). New York: Academic Press.
- Remez, R. E. (2003). Establishing and maintaining perceptual coherence: unimodal and multimodal evidence. *Journal of Phonetics, 31*(3-4), 293-304.
- Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101*(1), 129-156.
- Secord, W. A., Boyce, S. E., Donohue, J. S., Fox, R. A., & Shine, R. E. (2007). *Eliciting Sounds: Techniques and Strategies for Clinicians* (2nd ed.). Clifton Park, NY: Thomson Delmar Learning.
- Simpson, A. P. (1998). Accounting for the phonetics of German *r* without processes. *ZAS Working Papers in Linguistics, 11*, 91-104
- <http://www.ipds.uni-kiel.de/kjk/forschung/lautmuster.en.html>.
- Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging, 11*(2), 333-341.
- Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

- Tunley, A. (1999). *Coarticulatory influences of liquids on vowels in English*. University of Cambridge, Cambridge.
- Ulbricht, H. (1972). *Instrumentalphonetisch - auditive r-Untersuchungen im Deutschen*. Berlin: Akademie Verlag.
- Wagener, K. C., & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology*, 44(3), 144-156.
- West, P. (1999a). The extent of coarticulation of English liquids: An acoustic and articulatory study. In J. Ohala, J. Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (Eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences* (Vol. 3, pp. 1901-1904). Berkeley: University of California.
- West, P. (1999b). Perception of distributed coarticulatory properties of English / l / and / ɹ /. *Journal of Phonetics*, 27(4), 405-426.
- Zhou, X. H., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English / r /. *Journal of the Acoustical Society of America*, 123(6), 4466-4481.

## FIGURE CAPTIONS

Figure 1: Circles: mean differences in frequency (Hz), between /l/ and /r/ contexts measured at the midpoints of the periodic portion of the signal corresponding to the vowels in *be*, *could*, and *it*, and near the end of the diphthong in *might*, in the utterances *we heard it might be* \_\_, *we heard that it might be* \_\_, *we heard it could be* \_\_, *we heard that it could be* \_\_, each spoken with two different stress patterns by one male speaker of SSBE. \_\_ indicates one word of a minimal pair differing in /r/ vs /l/. Positive values mean /l/ > /r/. Left panels: unstressed *might* or *could*; right: stressed *might* or *could*. The size of the circles is proportional to the number of data points. The x-axis represents time relative to the start of the /r/ or /l/ (at 0 ms). The horizontal lines indicate the time range of the distribution of each vowel. (From Heid & Hawkins, 2000).

Figure 2: Percent target word correct for the four splicing conditions, l-base and l-word (ll), l-base and r-word (lr), r-base and r-word (rr), r-base and l-word (rl). A: native English speakers. B: all native German speakers. C: short-stay German ( $\leq 20$  months in UK). D: long-stay German ( $> 20$  months in UK). Bars are the standard error of the mean, adjusted to remove between-subject variance according to Loftus and Masson (1994).

Figure 3: Percent correct for four words relative to the splice point: +1 = target word; -1 = word immediately before the target word (and the splice point); -2, -3 = words earlier in the sentence. See text for further explanation. Bars are the standard error of the mean, adjusted to remove between-subject variance according to Loftus and Masson (1994). Where no bar is visible, the standard error was smaller than the physical symbol in the graph.

**Appendix 1. The 18 sentence pairs of Experiment 1.**

1. Ich habe heute eine Lüge / Rüge gesehen.
2. Ich habe heute einige Lanzen / Ranzen gesehen.
3. Ich habe heute eine Last / Rast gesehen.
4. Ich habe heute eine Kehle / Kehre gesehen.
5. Ich habe heute eine Aula / Aura gesehen.
6. Ich habe heute einige Kuhlen / Kuren gesehen.
7. Ich habe heute eine Linde / Rinde gesehen.
8. Ich habe heute eine Latte / Ratte gesehen.
9. Ich habe heute eine Leiche / Reiche gesehen.
10. Ich habe heute einige Gleise / Greise gesehen.
11. Ich habe heute eine Schlanke / Schranke gesehen.
12. Ich habe heute eine Klippe / Krippe gesehen.
13. Ich habe heute einige Flanken / Franken gesehen.
14. Ich habe heute eine Planke / Pranke gesehen.
15. Ich habe heute eine Liege / Riege gesehen.
16. Ich habe heute einige Gläser / Gräser gesehen.
17. Ich habe heute eine Laute / Raute gesehen.
18. Ich habe heute einige Länder / Ränder gesehen.



## Appendix 2. The 52 experimental sentence pairs of Experiment 2

Target words are in bold font. Words in italics (the mask) were visible on the computer screen while listeners responded. Underlined words are those scored for the analysis discussed in Figure 4, when intelligibility not only for the target words but also for three words before the splice point was considered.

### 1) word-onset l/r:

1. *Eddie says he's bought a **clock/crock** with his name on it.*
2. *I'm sure that many people can imagine the **clones/crones** have a bad name in the media.*
3. *For want of anything better, they ended up terming it the punitive **load/road** tax.*
4. *We're beginning to wonder why the dog **licks/ricks** his back so much.*
5. *The boys were searching for six **lambs/rams** on the farm.*
6. *We thought we could just distinguish some **lakes/rakes** outside.*
7. *She tied the boat up near the biggest **lock/rock**.*
8. *Bobby was disappointed that they couldn't save the damaged **leaf/reef**.*
9. *They don't want her to cause another nasty **clash/crash**.*
10. *Amy wondered why she wanted her to see the **flute/fruit** in the kitchen.*
11. *Esther said that she'd heard there were **clouds/crowds** on the day of the fair.*
12. *The boy was hoping he could adjust the **clutch/crutch**.*
13. *Emma was astounded to hear that he'd **bleached/breached** it.*
14. *I heard that the workers were fixing the **plough/prow** the other day.*
15. *Timothy thinks it's on the **blink/brink**.*
16. *The men were talking about the **plank/prank** in the harbour.*
17. *We were happy to see that they were **glowing/growing** in the dark.*
18. *He thought that some of the boys were **playing/praying** in the churchyard.*
19. *We don't want you to talk about the **splinter/sprinter** today.*
20. *By mere chance Fay could observe most of the **glaziers/graziers** in the competition.*

## 2) word -medial l/r:

21. *The paper says that according to Mr Colin/Corin Patterson this could be a new beginning.*
22. *We went to the modern art show to see Turner's painting of The Leaper/Reaper.\**
23. *She made the important point that aloes/arrows are cheap.*
24. *We sent the evidence to the Henley/ Henry Barker Company.*
25. *Jane misheard him. What he said he was doing was tiling/tiring.*
26. *Her forebears are thought to have moved to Delhi/ Derry six years ago.*
27. *So we became part of the Halley/Harry Davis folk band.*
28. *According to some sources, the business was set alight/aright by the new manager.*
29. *Jody wondered why they'd wanted to elect/erect the biggest one.*
30. *Tom asked why they were covered with molasses/morasses.*
31. *The workers over the way are supportive of the filing/firing system we've just bought.*
32. *They were concerned when they heard that it was hallowed/harrowed.*
33. *The teacher said that the best word was bestially/bestiary.*
34. *He went to the meeting just to hear the talk about Shelley/sherry.*
35. *She was astounded that he said we'd been believed/bereaved.*
36. *The Saturday papers said that the teller/terror terminated the voting.*
37. *We heard that some of the lighters/writers have been photo'd.\**
38. *They were good, so it's a pity they were belated/berated.*
39. *Some of the magazines were collected/corrected on Wednesday.*
40. *She wanted to name her baby Selina/Serena.*
41. *They thought that it might have been malicious/Mauritius.*
42. *It's a pity that the kids are bawling/boring.*
43. *What you should notice is that it's odourless/odorous.*
44. *Whenever she sees him she finds the pilot/pirate quite funny.*
45. *Abby wanted to study the palate/parrot on the counter.*

46. *We wondered* if his boss ever touched on **Allen's/Aaron's** work.
47. *We dug* the deep beds for the **mallow/marrow** in the garden.
48. *Peter went* over the way to the **Kelly/Kerry** Smith shop.
49. *Sadie hoped* to see the **milller/mirror** today.
50. *What* you should notice is whether the **belly/berry** turns pink now.
51. *She's not* said anything about **jelly/Jerry** for months now.
52. *She* was sitting there, just watching **telly/Terry**.

\* Target words in sentences 22 and 37 are in the word-medial group because, to produce high-quality splices, they were spliced at the beginning of the previous word (*the* in both cases). Thus, though not phonologically in the word-medial class, these stimuli were correctly classified phonetically, inasmuch as the vocoid portion of the signal immediately before the critical phoneme was included with the target word.

\*\* The length of the mask and location of the target word meant that only two preceding words could be scored in sentences 4 and 39.

Table 1: Means (SD) of the frequencies of F1-F4 in the schwas (underlined) in *heute* and *eine/einige* read in the context of l- and r-target words (see Appendix 1). The top row in parts A-C shows the mean duration (and SD) between the midpoint of the measured schwa and the onset of the r/l acoustic segment. A: Schwas in the carrier sentence when the critical /r/ or /l/ is word-initial. B: Schwas in the carrier when the critical /r/ or /l/ is C2 in a word-initial consonant cluster. C: Schwas in the carrier when the critical /r/ or /l/ is word medial. Data are averaged over ten speakers. The table also shows which sentences contributed to each computation (numbers according to Appendix 1). Bold font: l-r pairs that differed significantly in a paired-sample t-test. Table 1D shows size and direction of the difference in formant frequency that reached statistical significance, and the details of the paired-sample t-test.

A)

	<b>Word-initial singleton /r/ or /l/</b>							
	Singular (# 1, 3, 7, 8, 9, 15, 17)				Plural (# 2, 18)			
	<i>heute</i>		<i>Eine</i>		<i>heute</i>		<i>einige</i>	
	l	r	l	R	l	r	l	r
Duration from l/r in ms	268 (96)	275 (90)	106 (82)	105 (90)	250 (86)	255 (92)	92 (72)	99 (87)
F1	<b>570</b> <b>(67)</b>	<b>591 *</b> <b>(66)</b>	<b>426</b> <b>(85)</b>	<b>532 **</b> <b>(97)</b>	393 (84)	377 (50)	<b>402</b> <b>(43)</b>	<b>452 **</b> <b>(56)</b>
F2	<b>1684</b> <b>(181)</b>	<b>1650 *</b> <b>(148)</b>	<b>1895</b> <b>(189)</b>	<b>1699 **</b> <b>(175)</b>	2296 (252)	2297 (197)	<b>2018</b> <b>(261)</b>	<b>1839 **</b> <b>(211)</b>
F3	2648 (251)	2645 (259)	<b>2736</b> <b>(254)</b>	<b>2783 *</b> <b>(264)</b>	2847 (166)	2795 (219)	<b>2684</b> <b>(179)</b>	<b>2578 *</b> <b>(194)</b>
F4	3929 (342)	3916 (336)	<b>4008</b> <b>(342)</b>	<b>3800 **</b> <b>(303)</b>	3928 (305)	3924 (305)	3940 (381)	3820 (370)

B)

Word-initial /r/ or /l/ in a cluster with an preceding obstruent								
Singular (# 11, 12, 14)					Plural (10, 13, 16)			
<i>heute</i>			<i>Eine</i>		<i>heute</i>		<i>einige</i>	
	l	r	l	R	l	r	l	r
Duration from l/r in ms	243 (84)	254 (83)	96 (85)	95 (85)	283 (114)	296 (125)	116 (96)	120 (121)
F1	581 (73)	579 (71)	407 (74)	389 (93)	374 (49)	379 (52)	359 (32)	358 (34)
F2	1658 (139)	1660 (131)	<b>1894</b> <b>(243)</b>	<b>1846 **</b> <b>(230)</b>	2299 (239)	2316 (201)	<b>2048</b> <b>(288)</b>	<b>1980 **</b> <b>(281)</b>
F3	2638 (260)	2631 (254)	2696 (269)	2711 (316)	2846 (222)	2854 (210)	2596 (258)	2580 (280)
F4	3899 (339)	3925 (356)	3916 (385)	3852 (308)	3932 (274)	3940 (295)	3827 (309)	3832 (365)

C)

Word-medial /r/ or /l/								
Singular (# 4, 5)					Plural (# 6)			
<i>heute</i>			<i>eine</i>		<i>heute</i>		<i>einige</i>	
	l	r	l	r	L	r	l	r
Duration from l/r in ms	265 (96)	288 (95)	101 (84)	112 (98)	260 (62)	289 (59)	94 (40)	104 (48)
F1	581 (69)	604 (82)	521 (156)	546 (171)	378 (48)	374 (59)	345 (59)	349 (38)
F2	1653 (146)	1665 (159)	1918 (320)	1911 (328)	2287 (217)	2348 (171)	1932 (327)	1976 (199)
F3	2640 (236)	2649 (244)	2719 (260)	2768 (260)	2895 (244)	2645 (606)	2634 (380)	2573 (264)
F4	3885 (365)	3920 (312)	3869 (293)	3920 (299)	3995 (345)	3953 (267)	3916 (477)	3878 (423)

\*  $p \leq 0.05$ \*\*  $p \leq 0.005$

D)

		Formant	Direction of difference	Absolute difference (Hz)	df	t	p
<b>Word-initial singleton /r/ or /l/</b>							
<b>Singular</b>	<i>heute</i>	F1	l < r	21	69	2.82	0.006
		F2	l > r	34	69	2.38	0.020
	<i>eine</i>	F1	l < r	106	69	8.36	0.001
		F2	l > r	196	69	10.60	0.001
		F3	l < r	47	69	2.29	0.025
<b>Plural</b>	<i>einige</i>	F4	l > r	208	69	6.92	0.001
		F1	l < r	50	19	3.40	0.003
		F2	l > r	179	19	4.78	0.001
		F3	l > r	106	19	2.52	0.021
<b>Word-initial cluster /r/ or /l/</b>							
<b>Singular</b>	<i>eine</i>	F2	l > r	48	29	3.15	0.004
<b>Plural</b>	<i>einige</i>	F2	l > r	68	29	3.05	0.005

Table 2. Participants' age, gender, vocabulary score and length of stay in the UK and other English-speaking countries (ESC). M = mean, SD = standard deviation.

Group	Age M (SD)	Gender (m/f)	Mill Hill Vocabulary score M (SD)	Months in UK M (SD) (range)	Months in ESC M (SD) (range)
English	20.7 (3.0)	9/18	15.56 (1.63)		
German	29.0 (5.4)	7/20	13.26 (2.77)	29.26 (26.26) (3 – 105)	39.65 (38.81) (5 – 141)

Table 3. Age, gender, number of months spent in the UK, and number of months spent in other English-speaking countries (ESC) for each participant in the group of native German speakers.

Participant	Age	Gender	Months in UK	Months in other ESC
1	26	Male	3	12 (USA)
2	30	Male	5	0
3	26	Female	5	12 (USA)
4	24	Male	8	12 (USA)
5	25	Female	10	0
6	37	Female	10	0
7	22	Female	10	0
8	19	Female	10	0
9	23	Male	10	12 (USA)
10	27	Female	11	0
11	25	Female	12	0
12	30	Female	12	96 (USA)
13	40	Female	14	0
14	29	Female	20	0
15	26	Male	21	0
16	22	Female	22	0
17	30	Female	34	0
18	32	Female	36	0
19	34	Male	42	0

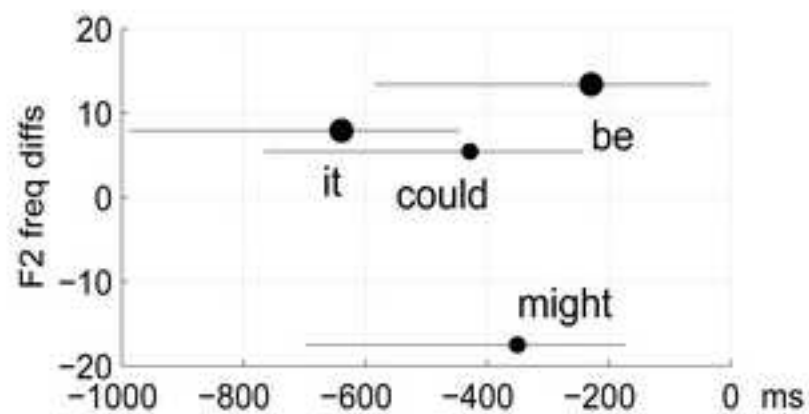
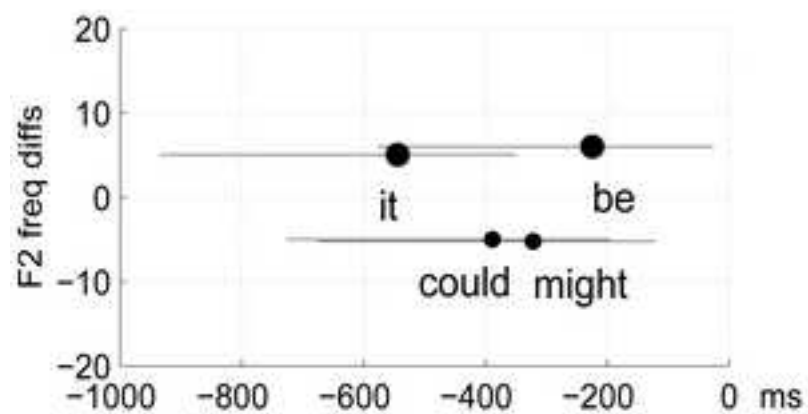
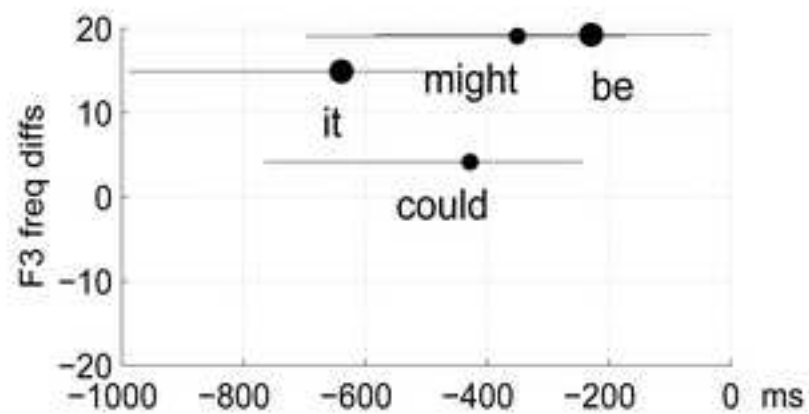
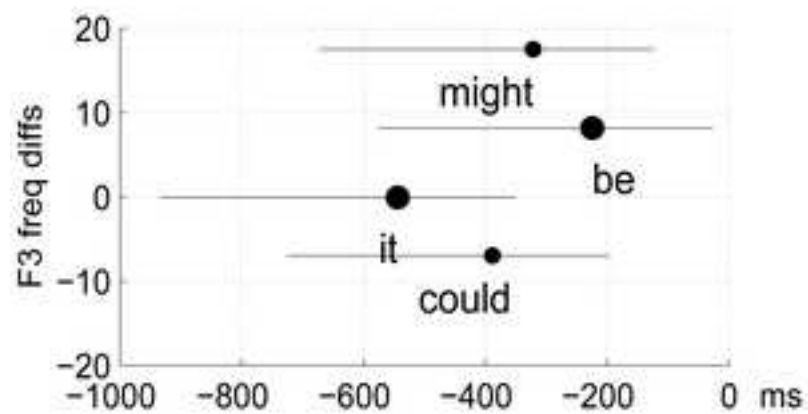
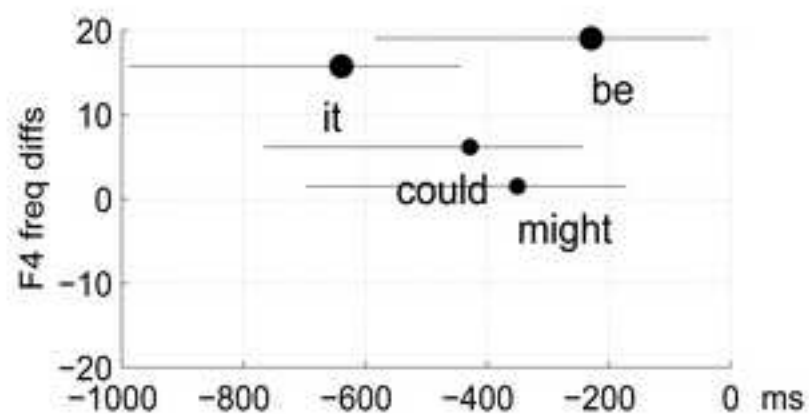
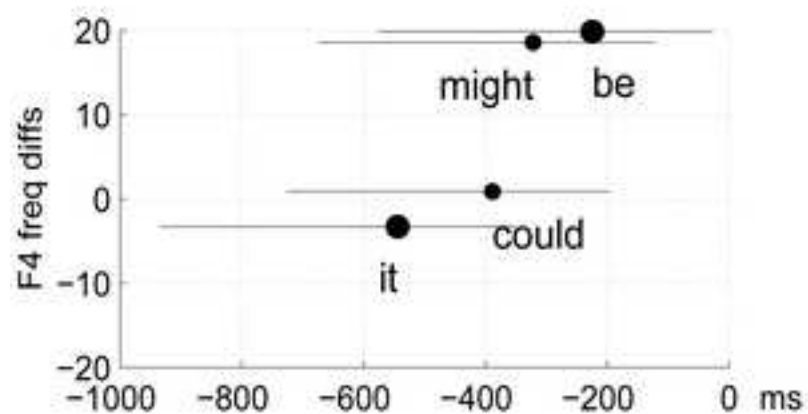


20	27	Female	42	0
21	35	Female	45	0
22	33	Female	45	0
23	38	Male	45	0
24	35	Female	60	60 (USA)
25	26	Female	69	0
26	34	Female	84	12 (USA)
27	28	Female	105	18 (Australia)

Table 4. Pearson product-moment correlation coefficients between intelligibility and target word frequency (BNC norms) (Leech et al., 2001) for English and German listeners.

splicing condition	English listeners N = 27	German listeners short-stay N = 14	German listeners long-stay N = 13
ll	0.301*	0.093	0.308*
lr	0.121	0.255	0.302*
rr	0.150	0.250	0.307*
rl	0.121	0.284*	0.284*

\*  $p < 0.05$



unstressed might/could

stressed might/could

