RHYTHMIC SEGMENTATION IN AUDITORY ILLUSIONS – EVIDENCE FROM CROSS-LINGUISTIC MONDEGREENS

Gerrit Kentner

Institut für Linguistik, Goethe-Universität Frankfurt
gerrit@lingua.uni-frankfurt.de

ABSTRACT

When confronted with non-native songs, listeners occasionally experience auditory illusions and perceive words in their native language although they are fully aware of the song lyrics being sung in a different language. We compiled two corpora with the original language of the song lyrics being English and the percept being either German or French. Against these two corpora, we tested the rhythmic segmentation hypothesis, specifically examining the cases of juncture misperceptions. The findings suggest that both German and French speakers use prominent syllables as anchors for segmentation, but they do so in language specific ways. For German listeners, prominent syllables signal the onset of lexical words. For French listeners, prominent syllables indicate phrase-finality. This cross-linguistic difference in boundary cueing corresponds with the specific role of prominent syllables in these languages, and makes a strong case for the concept of native listening in the context of sung speech.

Keywords: misperception, song lyrics, rhythmic segmentation, native listening

1. INTRODUCTION

Segmentation of the continuous speech stream into its component words is a task that listeners automatically perform not only when faced with spoken speech but also when confronted with sung speech.

The automaticity of segmentation in sung speech is vividly demonstrated by cross-linguistic mondegreens (misheard song lyrics) where listeners spontaneously perceive words or phrases in their native language even though they are fully aware that the song is actually being sung in a different language. This phenomenon has established itself as a popular meme and is a regular topic on radio shows where listeners report their misperceptions for comic effect. Given the automaticity of the auditory illusion [14], cross-linguistic mondegreens (also known as *soramimi*) can be considered an interesting test case for the study of speech perception in song (cf. [2]).

Previous work on speech segmentation suggests that listeners make use of heuristic segmentation procedures that are based on the experience with the structure of their native language [4]. One such heuristic is the metrical or rhythmic segmentation strategy (henceforth RSS) [3, 6]. According to the RSS, listeners use acoustically prominent (i.e. stressed or strong) syllables as anchors for speech segmentation but they do so in language-specific ways. Depending on the role of stress or prominence in the native prosodic system, prominent syllables may be more or less important for speech segmentation compared to other segmentation cues (e.g. phonotactics). Moreover, while in some languages, prominences indicate the beginnings of words or phrases, they may signal the end of comparable units in other languages.

In the case of German, a language with lexical stress and a preponderance of trochaic and dactylic lexical words in the native vocabulary [9], strong syllables are likely to be the initial syllables of lexical words. In contrast, weak or unstressed syllables typically represent grammatical words. Due to the strong propensity towards trochaic words (partly guaranteed by unstressed inflectional or derivational suffixes [16]), unstressed syllables that occur in lexical words are more likely to appear in non-initial positions (in spite of a comparatively rich system of unstressed prefixes).

French, in contrast, does not have clear lexical stress; rather, acoustically prominent syllables (typically those with a clear pitch excursion) have a demarcative function on the level of the phonological phrase rather than on the level of the word [12]. Phonological phrases roughly correspond to syntactic phrases (XPs) and obligatorily exhibit final prominence ('primary phrase final accent' [1], see also [12, 7]). A 'secondary' phrase-initial F0 rise is optional and not necessarily produced on the very first syllable of the phrase [12]. The initial F0-rise is a characteristic of, and possibly limited to, certain emphatic registers (e.g. political speeches) [8]. Thus, prominences should chiefly indicate phrase-finality in French.

Various studies attest the general validity of the

RSS across several languages using different research paradigms. Explicit evidence for the use of the RSS in the perception of song lyrics is, to the best of my knowledge, currently missing.

Most of the mondegreens we collected are more than one word in length. This offers the opportunity for misperceptions of word boundaries, so called juncture misperceptions [5]. We examine these juncture errors in two corpora of misheard song lyrics with the original language of the song being English and the percept being either German or French (henceforth English-German mondegreens and English-French mondegreens, respectively).

2. CORPUS STUDIES

2.1. Data and coding scheme

The data were gleaned from publicly available online platforms that archive radio programs in which listeners report misheard song lyrics. A considerable amount of these mondegreens is cross-linguistic. All mondegreens originate in pop songs of various genres that are accompanied by music. We were able to obtain 130 English-German mondegreens and 154 English-French mondegreens.[1] These were coded as follows. The mis-perceived part was aligned to the corresponding part of the original lyrics and the placement of word boundaries relative to the syllables was compared. Apart from the cases with coinciding word boundaries (36 English-German mondegreens, 51 English-French mondegreens), there are two types of possible boundary misperceptions: either a boundary insertion, i.e. a word boundary in the percept without corresponding word boundary in the original lyrics (1), or a boundary deletion, i.e. a single word in the percept spanning a word boundary in the corresponding original lyrics (2). Many mondegreens involve cases of both boundary insertions and deletions (3).

(1) Hope of de - li.ve   - rance *orig. lyrics*
     Hau   auf die | Le.ber | wurst *Germ. percept*
     "hit on the liver sausage"

(2) Let's   live it | up *orig. lyrics*
     Laisse les vé - los *Fr. percept*
     "Leave the bikes"

(3) They | al - ways send the   poor *orig. lyrics*
     Thé - o | et   ses deux porcs *Fr. percept*
     "Theo and his two pigs"

The concept of word boundary used here is coarsly phonological: contractions to monosyllables, such as Engl. *you're* or Fr. *d'la* ~ *de la* (Engl. '*of the*'),

were treated as a single words. However, cliticized function words contributing their own syllable (engl: *kinda* ~ *kind of*) were treated as separate words. Stems in compounds were counted as separate words, e.g. Germ. *Leber|wurst* (engl. '*liver sausage*'). Since the original lyrics and the percept are often not homophonous but rather distantly assonant, segmental detail, such as the affiliation of consonants to onsets or codas, was disregarded for the analysis.

Table 1 lists the number boundary misperceptions in the two corpora.

Table 1: Number of boundary misperceptions in the two corpora, broken down by type (insertion or deletion)

Percept	Insertions	Deletions	Total
German	50	87	137
French	75	99	174

In order to assess the validity of the RSS, the prominence of all syllables of the original (English) lyrics flanking the relevant word boundaries (inserted or deleted) were ascertained using an online dictionary of American English [13]. A syllable was marked as strong or stressed if it is quoted as having primary or secondary stress within the word it appears in. All other syllables were marked as weak or unstressed. Note that establishing the syllable prominence this way disregards its actual prominence in the song. Since, however, there does not seem to be a trivial way to assess the actual prominence of syllables in a song, the present approach seems to be an adequate approximation (any metric for actual syllable prominence in songs would certainly need to take into account, apart from lexical stress, absolute and relative acoustic measures on and around that syllable, let alone the characteristics of the instrumentation; complicating matters even more, the relative contribution of the various factors may be variable and song-specific).

2.2. Predictions

The general prediction of the RSS is that the distribution of word boundaries in inter-lingual auditory illusions based on non-native (here: English) song is dependent on the distribution of prominences in the lyrics. The specific dependency of boundary distribution and the distribution of prominences is subject to the role of prosodic prominence in the native linguistic system and thus differs between language groups.

2.2.1. English-German mondegreens

German is a language with lexical stress, i.e. basic prominences are assigned at the lexical level. Content words in German exhibit a strong tendency towards a trochaic pattern [9, 16]. Compared to German, the native vocabulary of English has a stronger tendency for lexical words to be stressed monosyllables [10]. Correspondingly, German listeners experiencing auditory illusions on the basis of English song lyrics are predicted to segment the sung input before strong syllables and these strong syllables then represent the initial syllables of lexical words in the percept. Moreover, German listeners are predicted to integrate unstressed syllables with preceding strong syllables into trochees; therefore, before unstressed syllables, boundaries that are present in the original lyrics are likely to be deleted.

As in English, grammatical words in German tend to be monosyllabic and unstressed. If boundaries are inserted before weak syllables, these weak syllables are likely to represent grammatical words.

2.2.2. English-French mondegreens

The French prosodic system assigns prominences at the phrasal level with the final syllable of every phrase carrying a boundary tone. The RSS would predict French listeners converting English song lyrics into French illusions to use the lexical prominences of the original English lyrics primarily as signal to a final boundary. That is, boundary insertions should preferably be placed after stressed syllables. Unstressed syllables, in contrast, are typically not phrase-final in French. Correspondingly, word boundaries after unstressed syllables in the original lyrics should be deleted in the French illusion. French prosodic phrases roughly correspond to syntactic phrases and syntactic phrases are likely to start in grammatical words (e.g. noun phrases, prepositional phrases, verb phrases in the periphrastic past tense (*passé composé*) typically start in function words: Det N, Prep Det N, Aux V). The English-French mondegreens should reflect this. If taken as phrase-final boundary tones, the prosodic prominences are likely to be followed by a (phrase-initial) grammatical word.

2.3. Results

2.3.1. English-German mondegreens

Boundary insertions occur nearly equally before weak syllables and before strong syllables, while boundary deletions are clearly more common, especially before weak syllables as compared to strong syllables, cf. Figure 1, left panel. A linear mixed model with binomial link function was applied to test the influence of following syllable prominence (strong vs weak) on the type of mis-segmentation (boundary deletion vs insertion); to account for the fact that cases of mis-segmentation from the same mondegreen are not independent, the song and the specific mondegreen were included as random factors. This model yields a significant main effect of syllable prominence (β=-1.072, SE=.387, z=-2.77, p=.0058), confirming the prediction that German listeners tend to insert boundaries before strong syllables and especially avoid word boundaries before weak syllables.

The right panel of Figure 1 shows all cases of boundary insertions broken down by the prominence of the following syllable. The plot suggests a clear dependence of word class and syllable prominence. That is, as predicted, boundaries inserted before strong syllables tend to produce lexical words while boundaries inserted before weak syllables produce grammatical words. A second linear mixed model probing whether the prominence of the syllable following a mis-inserted boundary reliably predicts the word class, confirms this dependence (β=2.539, SE=0.685, z=3.706, p=.0002).

Figure 1: Left panel: Mis-segmentations in English-German mondegreens broken down by type (deletion / insertion) and prominence of following syllable. Right panel: Boundary insertions broken down by type of following word (lexical / grammatical) and prominence of following syllable.
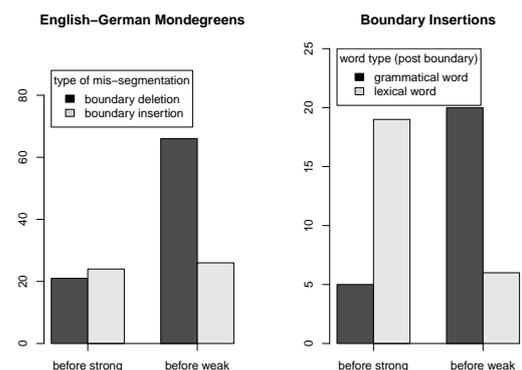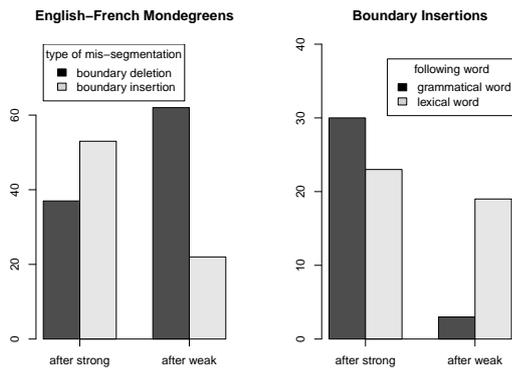

2.3.2. English-French mondegreens

As in the English-German mondegreens, there are more boundary deletions than insertions in the English-French mondegreens. Again, the distribution of boundary insertions and deletions is dependent on the position of strong syllables (cf. Figure

2, left panel). When compared to the original English lyrics, boundaries in the French percept are preferably inserted after strong syllables and deleted after weak syllables. A mixed logistic regression with song and mondegreen included as random effects confirms the significance of this interaction ($\beta=-1.396$, $SE=0.328$, $z=-4.257$, $p<.001$). The right panel of Figure 2 shows the boundary insertions broken down by the prominence of the preceding syllable and the word type of the following word. All in all, post-boundary words are more likely to be lexical words. However, in accordance with the predictions, when the inserted boundary is placed after a strong syllable, the following word is clearly more likely to be a grammatical word. The interaction between pre-boundary syllable prominence and post-boundary word class is highly significant ($\beta=-2.112$, $SE=0.68$, $z=-3.104$, $p=0.0019$).

Figure 2: Left panel: Mis-segmentations in English-French mondegreens broken down by type (deletion / insertion) and prominence of preceding syllable. Right panel: Boundary insertions broken down by type of following word (lexical / grammatical) and prominence of preceding syllable.

3. DISCUSSION

The analyses of the English-German and English-French mondegreens provide a strong case for the general validity of the RSS in non-native song perception. In both corpora, clear dependencies between the type of juncture misperception (insertion or deletion) and the prominence of the surrounding syllables were found. That is, both German and French native speakers use strong syllables as anchors for segmentation when confronted with English songs. Crucially however, as predicted, the specific anchoring of strong syllables differs according to the role of strong syllables in the native prosodic system. For German listeners, strong syllables signal onsets of lexical words, in accordance

with the preponderance of the trochaic foot structure in the German vocabulary. For the same reason, word boundaries before weak syllables are clearly avoided, as attested by the high proportion of boundary deletions in this context. Most likely, the overall bias for boundary deletions reflects the general tendency for lexical words in German to be trochaic compared to the preponderance of monosyllables in English [10]. Importantly, again in accordance with the RSS, the distribution of boundary insertions relative to surrounding syllable prominence predicts the word class of the post-boundary word, i.e. boundaries before strong syllables produce lexical words and boundaries before weak syllables produce grammatical words.

In contrast to German listeners, French listeners tend to posit boundaries not before but after strong syllables. Given the obligatory phrase-final accent of French, strong syllable are most likely not taken to correspond to lexical stress but to phrase-final prominence. Therefore, boundaries after strong syllables chiefly correspond to phrase boundaries, and only by implication to junctures between words. Since phrases preferably start in grammatical words, a grammatical word is predicted to follow a boundary inserted after a strong syllable – and this tendency is indeed confirmed in the mondegreen data.

Finally, a cautionary note about the nature of the data is in order. Since the mondegreens originate from songs accompanied with music, the misperceptions may well be based on the structure of the ambient music. Research on song perception has shown that linguistic and musical perception are strongly intertwined [11]. In general, strong positions in the music (on-beat notes) likely coincide with strong positions in the original lyrics (on-beat syllables) but this correlation is certainly not perfect. The present study does not consider this potential nuisance factor, so it remains unclear to what extent the juncture misperceptions originate from the text or from the tune (cf. [15] for a similar point). Moreover, it has to be noted that this natural experiment lacks the control over the materials that would be desirable. The English-German and English-French mondegreens stem from different songs and are therefore, strictly speaking, incomparable; however, given that the results neatly confirm the predictions derived from the prosodic systems of the two languages, it is highly unlikely that the differences found are specifically due to the different English source songs. In all probability, the mondegreens reflect the true difference between the prosodic systems of German and French and thus confirm the use of the native linguistic system in non-native song perception.

4. REFERENCES

[1] Astésano, C., Bard, E. G., Turk, A. 2007. Structural influences on initial accent placement in French. *Language and Speech* 50, 423–446.

[2] Beck, C., Kardatzki, B., Ethofer, T. 2014 Mondegreens and soramimi as a method to induce misperceptions of speech content – Influence of familiarity, wittiness, and language competence. *PloS One* 9.1, e84667.

[3] Cutler, A., 1990. Exploiting prosodic probabilities in speech segmentation. In: Altmann, G. T. M. (ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge: MIT Press, 105–121.

[4] Cutler, A. 2012. *Native listening: Language experience and the recognition of spoken words*. Cambridge: MIT Press.

[5] Cutler, A., Butterfield, S. 1992. Rhyhtmic cues to speech segmentation: Evidence from juncture misperceptions. *J. Mem. Lang.* 31, 218–236.

[6] Cutler, A., Norris, D. 1988. The role of strong syllables in segmentation for lexical access. *J. Exp. Psy: HPP* 14, 113–121.

[7] Di Christo, A., 1998. Intonation in French. In: Hirst, D., Di Christo, A. (eds.), *Intonation systems: A survey of twenty languages*. Cambridge University Press, 195–218.

[8] Di Christo, A. 2000. Vers une modélisation de l'accentuation en français. Deuxième partie: Le modèle. *Journal of French Language Studies*, 10, 27–44.

[9] Eisenberg, P. 1991. Syllabische Struktur und Wortakzent. *Zeitschrift für Sprachwissenschaft*, 10, 37–64.

[10] Fucks, W. 1955. *Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen*. Wiesbaden: Springer VS.

[11] Gordon, R. L., Schön, D., Magne, C., Astésano, C., Besson, M. 2010. Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PLoS One*, 5.3, e9889.

[12] Jun, S. A., Fougeron, C. 2000. A Phonological model of French intonation. In: Botinis, A. (ed.), *Intonation: Analysis, Modeling and Technology*. Dordrecht : Kluwer, 209–242.

[13] Merriam Webster online dictionary. http://www.merriam-webster.com/dictionary/. (retrieved January 2015.)

[14] Otake, T. 2007. Interlingual near homophonic words and phrases in L2 listening: Evidence from misheard song lyrics. Proc. 16th ICPhS. Saarbrücken. 777–780.

[15] Vitevitch, M. S. 2002. Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of the ear. *Language and Speech*, 45.4, 407-434.

[16] Wiese, R. 1996. *The phonology of German*. Oxford: Clarendon Press.