Bilingualism: Language and Cognition: page 1 of 12 © Cambridge University Press 2012 doi:10.1017/S1366728912000168

RESEARCH NOTE

Disentangling accent from comprehensibility*

PAVEL TROFIMOVICH
Centre for the Study of Learning and Performance, Concordia University
TALIA ISAACS
University of Bristol

(Received: November 6, 2011; final revision received: January 30, 2012; accepted: March 18, 2012)

The goal of this study was to determine which linguistic aspects of second language speech are related to accent and which to comprehensibility. To address this goal, 19 different speech measures in the oral productions of 40 native French speakers of English were examined in relation to accent and comprehensibility, as rated by 60 novice raters and three experienced teachers. Results showed that both constructs were associated with many speech measures, but that accent was uniquely related to aspects of phonology, including rhythm and segmental and syllable structure accuracy, while comprehensibility was chiefly linked to grammatical accuracy and lexical richness.

Keywords: accent, comprehensibility, second language speech learning

A recent article in *The New York Times* described an elementary school teacher in Arizona (an immigrant from northern Mexico) who has struggled to keep her job because of the perception that her English accent does not allow her to perform her job appropriately (Lacey, 2011). Like this teacher, many other people in North America may have also been confronted by their employers, inspected by "accent police" for the clarity of articulation, forced to take university acting classes, or referred to speech pathologists or accent reduction specialists (Lippi-Green, 2011; Munro, 2003). Apart from raising awareness of possible civil rights violations and exposing "accent reduction specialists" with dubious methods, such stories illustrate an important issue, namely, that many educators, researchers, policy makers, and members of the general public equate non-native speakers' accents with their ability to communicate effectively. The goal of this report is to examine whether accent and comprehensibility (one aspect of communicative effectiveness) are indeed distinct constructs.

One of the central challenges for scholars and practitioners interested in the development of spoken language in bilinguals and second language (L2) speakers stems from what Levis (2005, p. 370) describes as the tension between two "contradictory principles". The first, the "nativeness principle", holds that the goal of L2 learning is to help speakers acquire a nativelike accent and eradicate traces of their native language (L1). The second, the "intelligibility principle", emphasizes that L2 speakers should essentially strive to be understandable to their interlocutors, while acknowledging that a non-native accent does not necessarily preclude successful communication. In line with previous research, ACCENT refers to listeners' judgments of how closely the pronunciation of an utterance approaches that of a native speaker (Munro & Derwing, 1999). INTELLIGIBILITY is defined as listeners' actual understanding of L2 speech, most often measured by examining listeners' accuracy of orthographic transcriptions of L2 speech (Munro & Derwing, 1999). Intelligibility is distinguished from the companion dimension of COMPREHENSIBILITY, which denotes listeners' perceptions of understanding as measured by listeners' scalar ratings of how easily they understand speech (Munro & Derwing, 1999).

The assumption that a non-native accent does not necessarily impede successful communication has found empirical support in studies showing that accent and intelligibility, while related, are partially independent dimensions. For instance, L2 speakers judged to be heavily accented may still be completely intelligible, whereas unintelligible speakers are always rated as heavily accented (Derwing & Munro, 2009). Likewise, for at least some L2 listeners, there is a weak relationship between accentedness ratings and intelligibility scores (Munro & Derwing, 1999). Finally, listeners' judgments of L2 speech may be mediated by non-linguistic factors, such as listeners' attitudes toward a given non-native

* This research was made possible through grants from Social Sciences and Humanities Research Council of Canada, Fonds québécois de la recherche sur la société et la culture, and Sir James Lougheed Award of Distinction. We are grateful to Hyojin Song, Yvette Relkoff, Cassandre McLean Ikauno, Margaret Levey, Kathryn MacFadden-Willard, Fabrizio Stendardo, Garrett Byrne, and Joseph Hartfeil for their help with data analyses, and to Tracey Derwing and Murray Munro for sharing some of their testing materials. We also thank Ron Thomson, Carolyn Turner, Sarita Kennedy, and four anonymous *BLC* reviewers for their helpful input and feedback on the content of this manuscript.

Address for correspondence:
Pavel Trofimovich, Concordia University, 1455 de Maisonneuve Blvd. W., Montreal, Quebec, Canada, H3G 1M8
pavel.trofimovich@concordia.ca

speaking community, including stereotyped expectations about specific speech patterns that may not even be present in the speech signal (Kang & Rubin, 2009; Rubin, 1992). Clearly, a listener's detection of an L2 accent does not always involve a lack of understanding.

Apart from these findings, little research has explored which linguistic features of speech are most crucial for intelligibility and which, while noticeable or irritating, merely contribute to the perception of an accent. One reason for this gap is that researchers have often considered only some variables as possible influences on accent and intelligibility. For example, Anderson-Hsieh, Johnson and Koehler (1992) found a link between several measures of phonology (segment, prosody, and syllable structure accuracy) and listener ratings of L2 speech. However, accent and intelligibility were conflated in the rating scale they used. Hahn (2004) showed that native listeners processed L2 speech with correct primary stress placement faster and more accurately than they did in incorrect or missing stress conditions, underscoring the importance of stress in contributing to listener understanding. In a study focusing on fluency, Derwing, Rossiter, Munro and Thomson (2004) showed that several temporal measures (pausing, articulation rate) were related to listener fluency judgments and, through examining statistical relationships between holistic ratings, suggested that these measures likely also affect comprehensibility. Similarly, only a handful of studies have explored morphosyntactic accuracy as an influence on speech ratings, citing an association between grammar ratings and listeners' perceptions of L2 speech (Fayer & Krasinski, 1987; Munro & Derwing, 1999; Varonis & Gass, 1982). And, to our knowledge, no study has explored how discourse factors, such as L2 speakers' narrative structure or use of cohesive devices, affect listener evaluations of L2 speech.

Another reason for the lack of empirical evidence about linguistic influences on accent and intelligibility is that different measures of L2 speech – including accent, intelligibility, and the companion dimension of comprehensibility – have rarely been examined in the same study in relation to several linguistic variables (see Munro & Derwing, 1999, for a rare exception). For example, in an investigation of Danish schoolchildren's L2 English speech, Albrechtsen, Henriksen and Færch (1980) explored the relationship between eight linguistic variables (lexical, syntactic, morphological, and segmental accuracy, intonation measures, speech rate, hesitation phenomena, and communication strategies) and English listeners' comprehensibility ratings. Only the measure of communication strategies was associated with comprehensibility, such that speech samples that included extensive use of particular strategies (e.g., language switch) were deemed difficult to understand. Although these findings provide insights into the linguistic dimensions that feed into comprehensibility, at least with respect to the tasks and participants targeted in that study, it is unclear whether these dimensions contribute to accent.

As the preceding discussion suggests, there is currently little understanding of the linguistic dimensions that are relevant to accent and those that contribute to intelligibility or comprehensibility. Our goal was therefore to extend previous research by examining 19 different speech measures (drawn from the domains of phonology, fluency, lexis, grammar, and discourse) in the speech of 40 native French speakers of English in relation to both ACCENT and COMPREHENSIBILITY, as rated by 60 novice raters and three experienced language teachers.[1]

The choice of comprehensibility as a measure of understanding, as opposed to more objective measures of intelligibility (see Isaacs, 2008), was motivated first by practical considerations. Although intelligibility is defined as listeners' actual understanding of L2 speech (Munro & Derwing, 1999), it is often used synonymously with comprehensibility to denote more generally listeners' ability to understand L2 speech (Levis, 2006). It is in this broader sense that intelligibility and comprehensibility have been referred to as the rightful goal of L2 teaching (Munro & Derwing, 1999) and, by implication, of L2 assessment (Levis, 2006). In the narrower sense, the distinction between intelligibility and comprehensibility relates to the way these constructs have been operationalized. In high-stakes assessment contexts, measuring listeners' understanding of L2 speech using oral proficiency scales is by far more common than eliciting listeners' transcriptions of each performance sample. Although several L2 oral proficiency scales make use of the term "intelligibility" in their rating band descriptors (e.g., TOEFL iBT, IELTS), the use of a scale to elicit listener ratings suggests that it is, in fact, narrowly-defined comprehensibility that is being used as the operational definition in these scales. Thus, comprehensibility (hereafter used in its narrow sense) reflects a typical and more practical approach to assessing broadly-defined intelligibility in oral proficiency scales and in many bilingual assessment settings.

In a recent study, Lev-Ari and Keysar (2010) provided another compelling reason for focusing on comprehensibility as a measure of understanding. These researchers showed that native listeners tend to perceive statements as being less truthful when spoken by accented L2 speakers, and attributed this finding to listeners' reduced processing fluency. Processing fluency, which refers to individuals' subjective experience of ease or difficulty in processing information on a cognitive task (Oppenheimer, 2008),

[1] The term "grammar" is used here to differentiate morphology and syntax from other aspects of language (e.g., phonology, lexis); it does not imply that these other aspects of language lack structural properties.

affects how information is judged, with the idea that, for example, easier to process information is perceived as more truthful, familiar, pleasant, or more distinct than information that is harder to process (Reber & Schwarz, 1999; Whittlesea, 1993). Clearly, when it comes to L2 speech, processing fluency (as defined by cognitive psychologists) would encompass at least some aspects of comprehensibility, which denotes listeners' subjective experience of the ease or difficulty of understanding speech. Thus, it is essential to identify which linguistic dimensions make L2 speech less comprehensible, thereby decreasing processing fluency for the listener, and which merely contribute to the perception of a foreign accent. Therefore, being able to disentangle accent from comprehensibility not only can help refine the assessment of bilingual and L2 speaking ability, but also can clarify the linguistic influences that feed into negative stereotyping based on accent (Lippi-Green, 2011) and that lead to the perception of less comprehensible speech as being less truthful (Lev-Ari & Keysar, 2010).

This study employed a sequential mixed-methods design (Creswell & Plano-Clark, 2011), with qualitative data used to complement and explain quantitative findings. The 60 novice raters' accentedness and comprehensibility ratings were first analyzed to identify the linguistic dimensions of L2 speech that show most statistically robust associations with accentedness and comprehensibility. The three experienced teachers' introspective reports were examined next, to determine whether the statistically most robust associations also featured among the linguistic factors that experienced teachers consider when rating L2 speech. Because "accent" and "comprehensibility" are complex holistic constructs that are not directly observable, multiple sources of evidence were necessary to examine whether unique components of these constructs could be identified. Asking both experienced teachers and novice raters to evaluate L2 speech also helped offset possible rater biases in each rater group. Indeed, raters that vary in amount of skill (e.g., musical ability) or expertise (e.g., teaching experience) might rate speech differently, exercising various degrees of severity or leniency in their decisions (Cartier, 1975; Isaacs & Trofimovich, 2010; Schairer, 1992). The overarching question was to determine which linguistic aspects of L2 speech are most strongly related to accent and which to comprehensibility.

Method

Participants

The target speech samples were elicited from 40 native French speakers of L2 English (27 females, 13 males) from Quebec, Canada (M_{age} = 35.6 years, range = 28–61 years) as part of an earlier study on L2 phonological learning (Trofimovich, Gatbonton & Segalowitz, 2007). All speakers, with the exception of two early French–English bilinguals, started learning English in elementary school (M_{age} = 9.3 years) in 45-minute weekly ESL classes and received up to three hours per week of subsequent ESL instruction. At the time of the study, the speakers estimated using English to varying degrees (0–70% of the time daily) and reported a range of English speaking, listening, reading, and writing abilities (a 1–9 proficiency scale, where 1 = "extremely poor", 9 = "extremely proficient").

Several additional measures of the speakers' L2 speaking ability were derived from the speech data collected as part of the earlier project (Trofimovich et al., 2007). These measures came from a 440-word English read-aloud task recorded onto a computer using a Plantronics (DSP-300) microphone. The recordings were subsequently evaluated by 10 English native-speaking expert raters for the accuracy of English /ð/ (as in *brother*), a difficult consonant for French speakers (0 = "does not sound like a good English /ð/", 1 = "sounds like a good English /ð/"). The recordings were also presented to five expert listeners (M_{age} = 38.2 years; all exposed to English from birth) to assess the degree of accent (a 1–9 accent scale, where 1 = "heavily accented", 9 = "not accented at all"). In addition, a measure of articulation rate (syllables/second) was obtained from the recordings, computed as the total number of syllables articulated (including repetitions, hesitations) over the total duration of the sample. The obtained average scores ranged between 7% and 99% correct for English /ð/ accuracy, between 1.8 and 9.0 for accent ratings, and between 0.4 and 3.4 syllables/second for articulation rate. The speakers thus represented different pronunciation ability levels, from beginner to advanced.

The speakers' extemporaneous speech was elicited using an eight-frame picture story depicting two travellers who bumped into each other on a busy street corner and accidentally exchanged the identical suitcases they were carrying, only realizing their mistake thereafter (Derwing, Munro & Thomson, 2008). The speakers, who were tested individually, first studied the picture story for about one minute and then recorded their narrative directly onto a computer (using a Plantronics DSP-300 microphone). The stories ranged in duration between 26.4 s and 322.8 s. To keep the content of the story consistent across speakers, the first few seconds of the stories were excised from the recordings by removing all dysfluencies (e.g., false starts) from the beginning of the story and by using natural pauses to demarcate the end of each excerpt. This procedure is consistent with previous ratings of speech samples from the same task (see Munro & Derwing, 1999) and with research showing that listeners make reliable judgments based on short samples (Munro, Derwing & Burgess, 2010). These excerpts were

then normalized for peak intensity and randomized for presentation to raters. The final samples (23–36 s in length) were orthographically transcribed and verified for accuracy by another transcriber.

Speech measures

The 40 speech samples were analyzed for 19 different measures, in order to capture as many linguistic variables as raters might use to arrive at their judgments of L2 accentedness and comprehensibility.[2] The measures roughly fell into four distinct categories (phonology, fluency, lexis/grammar, discourse), all based on prior research linking some of these measures to different aspects of speech, including accent, comprehensibility, and intelligibility (Albrechtsen et al., 1980; Anderson-Hsieh et al., 1992; Derwing et al., 2004; Fayer & Krasinski, 1987; Munro & Derwing, 1999; Varonis & Gass, 1982).

Phonology

This category included a mix of measures at the segmental level (individual vowels and consonants) and suprasegmental level (syllables, words, and phrases).

(1) Segmental errors: the number of phonemic (e.g., *think* spoken as *tink*) substitutions divided by the total number of segments articulated.

(2) Syllable structure errors: the total number of vowel and consonant epenthesis (insertion) and elision (deletion) errors (e.g., *holiday* spoken without the initial /h/) over the total number of syllables articulated.

(3) Word stress errors: the total number of instances of misplaced or missing primary stress in polysyllabic words (e.g., *SUIT-case* spoken as *suit-CASE*) divided by the total number of polysyllabic words produced.

(4) Rhythm, defined as vowel reduction ratio and used as a measure of English stress timing: the number of correctly reduced syllables over the total number of obligatory vowel reduction contexts in both polysyllabic words and function words (e.g., *a MAN and a WOman aRRIVES at the SAME TIME* contains 7 obligatory contexts, all in lowercase letters; the speaker pronounced "woman" as *wo-MAN* and, thus, produced six correct vowel reductions).

(5) Pitch contour, as a measure of intonation accuracy (Wennerstrom, 2001): the number of correct pitch patterns at the end of phrases (i.e., syntactic boundaries) over the total number of instances where pitch patterns are expected, as signaled by pre-boundary lengthening (e.g., the sentence *Once upon a time* [level tone] *a man and a woman were walking on the sidewalk* [falling tone] has two correct pitch patterns).

(6) Pitch range, as a measure of pitch breadth: the difference between the highest and lowest fundamental frequency (F0) values, as measured by using the *Praat* pitch tracker function (Boersma & Weenink, 2010). Inspired by Wennerstrom's (2001) paratone measure, this measure was expressed in absolute terms for each speech sample and was used to capture changes in pitch used by individual speakers to convey additional meaning and emphasis, with the idea that a narrower pitch range describes flat, monotone voices while a wider pitch range describes animated voices.

Fluency

This category comprised several measures of fluency meant to describe the speech samples in terms of dysfluencies often present in L2 speech.

(7) Total number of filled (non-lexical) pauses such as uh and um (e.g., *The first picture uh* [one filled pause] *we can see a town*).

(8) Total number of unfilled (silent) pauses (e.g., *In the first picture* [unfilled pause] *I I see buildings* [unfilled pause] *a a lot of buildings*). Following Derwing et al. (2004), only filled or unfilled pauses of 400 ms or longer were counted for the pause-based measures.

(9) Pause errors: the number of inappropriately produced filled and unfilled pauses (i.e., inside clauses and not at syntactic boundaries, where pauses would be expected), divided by the total number of pauses produced (e.g., *There was two* [unfilled pause] *people who uh* [filled pause] *were on on on* [unfilled pause] *business trip* [unfilled pause] *and were staying on the 11th floor*). Unlike the

[2] A reviewer questioned our decision to only examine L2 speech samples in this study. First, the phenomenon of interest here is L2 speakers' productions, not native speaker speech. This is because in many testing settings, L2 and bilingual speakers are not explicitly compared to native speakers but are instead evaluated against a particular criterion of L2 proficiency. Thus, the use of native speaker samples would have been incompatible with typical L2 assessment practices. Second, although accent is defined in reference to a native speaker, listeners generally have an internalized notion of what constitutes accented speech and do not require explicit comparisons with a native speaker performing the same task (e.g., Southwood & Flege, 1999). However, in the case of comprehensibility, there is much more scope for rater interpretation (Isaacs & Thomson, in press). The inclusion of native speaker speech might have encouraged the raters to reserve the high end of the comprehensibility scale for native speakers, although many accented L2 speakers are perfectly intelligible/comprehensible (Derwing & Munro, 2009). Finally, the obtained high interrater reliability (Cronbach $\alpha = .99$) suggests that our raters were consistent in their decisions in the absence of native speaker samples.

previous two pause counts, this measure captured the relationship between fluency and sentence structure.

(10) Repetitions/self-corrections: the sum of all immediately repeated and self-corrected words (e.g., *We are in New York city with a big big big* [repeated] *bag big* [self-corrected] *big* [repeated] *and high house*) over the total number of words produced.

(11) Articulation rate, defined as "pruned" syllables per second: the total number of syllables produced excluding dysfluencies (e.g., filled pauses, repetitions, self-corrections, false starts), calculated over the total duration of the speech sample.

(12) Mean length of run (MLR): the mean number of syllables produced between two adjacent filled or unfilled pauses.

Lexis/grammar

This category included several measures of grammatical and lexical accuracy in the speech samples.

(13) Grammatical errors: the number of words with at least one morphosyntactic error divided by the total word count. This included sentence structure, morphological, or syntactic errors (e.g., *They took back their uh suitcase and go to their place* contained one plural agreement error and one past-tense error).

(14) Lexical errors: the number of incorrectly used lexical expressions, including phonetically similar but semantically inappropriate words (e.g., *tied* instead of *tie*), false cognates (e.g., *quit* instead of *left*), imprecise vocabulary choice (e.g., *enter in contact* instead of *bump into*), incorrectly used lexical expressions (e.g., *wallet* instead of *suitcase*), and L1 intrusions (e.g., *malette* [for suitcase]), over the total number of words produced.

(15) Token frequency: the total number of words produced.

(16) Type frequency: the total number of unique words produced. Type and token counts were computed separately using the online *Vocabprofile* program (Cobb, 2000). To correct for differences in sample length, the raw counts for both token and type frequencies were divided by the total duration of the sample.

Discourse

This category involved several discourse-level measures, in an attempt to describe the speakers' storytelling strategies which may feature in raters' accent and comprehensibility judgments.

(17) Story cohesion (Martin & Rose, 2003): the number of adverbials used as cohesive devices (e.g., *suddenly*, *but*, *hopefully*) that help situate the listener in the story by establishing links between storytelling elements, propelling the storyline forward, or revealing the storyteller's attitude. To correct for differences in sample length, this and the remaining discourse measures were normalized by dividing frequency counts by the total duration of the sample.

(18) Story breadth (Stein & Glenn, 1979): the number of distinct propositions or storytelling elements (predicate + another argument) in a speech sample. Distinct proposition categories include setting (e.g., *In a big city, two person walk on the sidewalk*), initiating event (e.g., *when they were above to arrive at the corner*), attempt (e.g., *they bang each other on the head*), direct consequence (e.g., *they pick up their suitcase and continue on their way*), and reaction (e.g., *they were shocked and they were all dizzy*).

(19) Story depth: the number of different proposition categories in a speech sample (e.g., setting, attempt, reaction), based on the idea that a speech sample that features only the setting may be poorer in discourse structure than a sample that first focuses on setting the scene and then describes the events and consequences.

The 19 measures were first coded by a trained coder; then, another trained coder re-coded 40% of the speech samples for each of the 19 measures. Both coders were instructed to evaluate speech samples for each measure based on their native speaker intuition (e.g., decide whether the intonation contour was correct based on how they themselves would produce it if they were a speaker). As with all perceptual judgments, such judgments include a certain degree of subjectivity and may not reflect the variability found in spoken language. Nevertheless, high intercoder agreement was obtained, with intraclass correlations exceeding .90 for all measures except lexical error ratio (.85), suggesting that coding decisions were internally consistent.

Procedure

The speech samples were evaluated by 60 novice raters, native English-speaking undergraduate students (34 females, 26 males) from a variety of non-linguistic disciplines (e.g., physiology, music, sociology) at a Canadian English-medium university. The raters ($M_{age} = 20.7$ years, range = 19–25 years), who were from

monolingual homes in Canada (29) and the United States (31), reported English as their main language (used over 90% of the time daily) and indicated low familiarity with French. No rater reported hearing problems or hearing-related disorders. The raters evaluated the 40 speech samples individually using a Koss R/80 headset connected to a computer in a quiet office. After familiarizing themselves with the picture sequence, they listened to each story in randomized order and assigned ratings using separate nine-point Likert-type scales for accentedness (1 = "heavily accented", 9 = "not accented at all") and comprehensibility (1 = "hard to understand", 9 = "easy to understand"). Nine-point numerical scales have been used extensively in L2 pronunciation research (Derwing & Munro, 2009) following evidence from the accent scaling literature that at least nine levels are necessary to capture the magnitude of ratings that raters may detect when judging widely variable L2 speech samples in terms of pronunciation ability (Southwood & Flege, 1999).

Three additional experienced raters were then recruited to generate in-depth input on the aspects of L2 speech that listeners consider when judging L2 accent and comprehensibility. The raters were native English-speaking ESL teachers (2 female, 1 male) with graduate degrees in TESL, 10–12 years of classroom teaching experience, but no training in phonetics/phonology or assessment. Originally from English-speaking Canada, the teachers had resided in Quebec 8–24 years and had extensive experience teaching ESL to adult L1 French speakers. The teachers, who were tested individually, first studied the picture prompt and completed two sample ratings. To clarify initial understanding of the constructs, they were told that accentedness refers to "how different the speaker sounds from a native speaker of North American English" while comprehensibility denotes "how easy the speaker is to understand".[3] They then listened to the 40 speech samples in randomized order via a Koss R/80 headset. After listening to each sample, with multiple listenings permitted, they paused the recording to mark their accentedness and comprehensibility ratings (presented in that order) in an electronic response sheet using the nine-point accentedness and comprehensibility scales described above. They then described the aspects of speech that they attended to when scoring by typing their impressions into separate preformatted boxes for accentedness and comprehensibility. At the end of the session, the teachers were given a list of several factors describing the 19 target speech measures in lay terms (e.g., lexical errors, story cohesion, natural sounding rhythm) and were asked to select those that had most affected their ratings.

[3] A reviewer suggested that using "North American English" as a reference point may contribute unwanted variance to listener ratings, possibly due to listeners' sensitivity to regional dialectal differences of English. Although existing research is unclear as to how sensitive native speakers are to dialectal differences, with some studies reporting low dialect classification rates (e.g., Clopper & Pisoni, 2004), there is some evidence that listeners may be influenced by the presence of non-standard dialect features in the speech they rate (Robinson & Stockman, 2009). While these concerns highlight an interesting focus of future investigation, they are of little threat to the validity of this dataset because none of the 40 Quebec French speakers, based on our judgment, produced any perceptible regional dialect features outside of what would be expected in Canadian English (e.g., the low-back merger; see Labov, Ash & Boberg, 2006).

Results

Novice raters

The 60 novice raters were overall consistent in their ratings (Cronbach $\alpha_{acc.} = .99$, $\alpha_{comp.} = .99$). Therefore, for all further analyses, their accentedness and comprehensibility ratings were averaged to derive single mean ratings for each of the 40 speakers. Pearson correlations were then computed to examine the strength of associations between these mean ratings and the 19 analyzed speech measures. As Table 1 shows, accentedness and comprehensibility ratings, which were strongly correlated with each other ($r = .90$), showed medium-to-strong correlations ($r > .60$) with eight measures across the four conceptual categories: phonology (word stress, rhythm, segmental errors), fluency (MLR), lexis/grammar (type frequency, token frequency, grammatical accuracy), and discourse (story breadth). Six of these measures were common to accentedness and comprehensibility (word stress, rhythm, MLR, type frequency, token frequency, story breadth). The remaining two measures were unique to each dimension: segmental errors correlated with accentedness while grammatical accuracy correlated with comprehensibility. Weaker correlations were obtained for all remaining measures, with the exception of pitch range, where no statistical association was detected.

Regression analyses were conducted next to determine the combination of L2 speech measures that best account for the variance in the accentedness and comprehensibility scores assigned by the novice raters. Collinearity diagnostics, performed prior to conducting a separate regression for each rated measure (see Table 2), revealed that token frequency, MLR, and story breadth were all strongly associated with type frequency ($r \geq .75$), which alone accounted for 56–92% of variance in these measures. Therefore, of the seven measures most strongly correlated with accentedness and comprehensibility ($r > .60$), four were retained for inclusion as predictors for accentedness (word stress, rhythm, type frequency, segmental errors) and four for comprehensibility (word stress, rhythm, type frequency, grammatical accuracy).

Initial regression analyses employed the maximum R^2 improvement technique (SAS Institute, 2004), a procedure which uses R^2 (effect size) to identify the model with

Table 1. *Pearson Correlation Coefficients between L2 speech measures and 60 novice raters' scalar judgments of L2 accentedness and comprehensibility.*

Speech measure	Accentedness	Comprehensibility
Word stress errors	−.78**	−.76**
Rhythm (vowel reduction ratio)	.74**	.74**
Story breadth	.67**	.71**
Type frequency	.62**	.78**
Mean length of run (MLR)	.62**	.71**
Token frequency	.61**	.77**
Segmental errors	−.60**	−.54**
Grammatical accuracy	−.53**	−.63**
Pitch contour	.49**	.57**
Repetitions/self-corrections	−.48**	−.57**
Pause errors	−.46**	−.58**
Story depth	.44**	.42**
Lexical errors	−.41**	−.52**
Syllable structure errors	−.40*	−.37*
Total filled pauses	−.39*	−.45**
Articulation rate	.39*	.35*
Story cohesion	.32*	.50**
Total unfilled pauses	−.12	−.32*
Pitch range	−.02	−.07

*$p < .05$, **$p < .01$, two-tailed

the largest increase in R^2, as each successive predictor is added, until the R^2 for the full model is given. This procedure allows for comparisons of the models with the largest effect size for any possible number of predictors. The best-fitting model was chosen based on the C_p statistic (total squared error for a model with p variables), as suggested by Mallows (1964). For accentedness, the best model involved two variables, with word stress and rhythm as significant predictors, $R^2 = .76$, *Adj.* $R^2 = .74$, $C_p = 1.87$, $p < .0001$. For comprehensibility, the best solution was a three-variable model with type frequency, word stress, and grammatical accuracy as predictors, $R^2 = .86$, *Adj.* $R^2 = .85$, $C_p = 4.76$, $p < .0001$, with no evidence of collinearity in either model ($VIF < 2.81$). To confirm model selection, parallel stepwise regression analyses were run with the original four variables used as predictors. Stepwise regression employs significant F-values associated with each predictor (as opposed to R^2) as the criterion for variable selection ($\alpha = .05$). This procedure yielded the same regression models for both accentedness and comprehensibility as the maximum R^2 improvement technique. Summary statistics for both models are shown in Table 3.

Table 3. *Summary of stepwise multiple regression analyses for speech measures as predictors of accentedness and comprehensibility.*

Variable	R^2	ΔR^2	F	df	p
Accentedness					
Rhythm	.63	.63	60.27	1,36	.0001
Word stress	.76	.13	19.41	1,35	.0001
Comprehensibility					
Type frequency	.64	.64	64.39	1,36	.0001
Word Stress	.80	.16	29.00	1,35	.0001
Grammatical accuracy	.86	.06	14.18	1,34	.0006

Experienced teachers

Introspective reports were examined next to determine if in evaluating L2 speech, experienced teachers actually consider those factors that statistically emerged as

Table 2. *Intercorrelations among the speech measures most strongly correlated with accentedness and comprehensibility.*

Speech measures	1	2	3	4	5	6	7	8
1. Word stress errors	–							
2. Rhythm	–.63**	–						
3. Story breadth	–.54**	.60**	–					
4. Type frequency	–.55**	.72**	.75**	–				
5. Mean length of run	–.52**	.63**	.67**	.88**	–			
6. Token frequency	–.50**	.72**	.73**	.96**	.86**	–		
7. Segmental errors	.46**	–.58**	–.32*	–.44**	–.44**	–.43**	–	
8. Grammatical accuracy	.45**	–.37**	–.36**	–.45**	–.47**	–.46**	.16	–

*$p < .05$, **$p < .01$, two-tailed

Table 4. *Frequency of coded categories for accentedness and comprehensibility (raw number and percentage) from teacher reports.*

	Accentedness			Comprehensibility		
Coded category	*n*	%	Teachers	*n*	%	Teachers
Vowels and consonants	95	27	T1, T2, T3			
Syllables	52	15	T1, T2, T3			
Sounds nativelike/non-nativelike	42	12	T1, T2, T3			
Rhythm	7	2	T1, T3			
Intonation	24	7	T1, T2, T3	4	1	T3
Inadequate words or information produced	6	2	T1, T2, T3	6	2	T1, T3
Word stress	13	4	T1, T2, T3	6	2	T3
Accent/pronunciation (general comment)	22	6	T1, T2, T3	20	7	T1, T3
Fluency	34	10	T1, T2, T3	29	10	T1,T2,T3
Vocabulary	12	3	T1, T2	38	14	T1,T2,T3
Grammar	30	9	T1, T2, T3	45	16	T1,T2,T3
Hard/easy to understand (general comment)	12	3	T1, T2	73	27	T1, T2, T3
Anyone can understand regardless of background				6	2	T2
Storytelling elements and cohesion				26	9	T1
Need to be a teacher, know the context, or have exposure to French to understand				29	10	T2
Total	349	100		282	100	

Note: T1, T2, T3 (Teachers 1, 2, 3) in the Teachers columns indicate whether the relevant categories were present in each of the experienced teachers' reports.

strongest predictors of accent and comprehensibility. The analysis of the teachers' typed reports was based on a 15-category coding scheme, with the 19 speech measures serving as the starting point. The raters' descriptive comments were first thematically coded according to a larger number of categories, then re-coded to eliminate overlapping ones (e.g., "L1 intrusions", "L1-influenced lexical items", and "odd lexical choice" were combined under "vocabulary"). Finally, all coded comments were "quantitized" by tabulating frequency counts for each category (Teddlie & Tashakkori, 2009). Following initial coding, 40% of the data were re-coded by a second blind coder. Exact intercoder agreement reached 96%, and all cases of disagreement were resolved through discussion. Table 4 shows the frequencies of the resulting coded categories.

As can be seen from Table 4, four categories uniquely distinguished accent from comprehensibility, with all categories specific to the dimension of phonology (i.e., vowels and consonants, syllables, sounding nativelike, and rhythm). Nearly a third of all comments for accentedness (27%) centered around speakers' pronunciation difficulties at the segmental level (e.g., "d sound for th, vowel in *city*", "his vowel sounds influenced my rating", "non-aspirated beginning 't'"), which included phonemic errors and phonetic detail. An additional 27% of the comments pertained to syllable-structure errors (e.g., "missing word endings", "pronounced the final 'e' when not present") and general comments about L2 speakers sounding nativelike (e.g., "not much accent at all", "had almost an American twang to the speech").

In contrast, two of the three teachers produced comments that focused on aspects of speech pertaining solely to comprehensibility; these dealt with discourse structure of the narratives (9%) and comments about the role of listener background characteristics or knowledge of context in evaluating comprehensibility (12%). However, the teachers most frequently commented on how easy or difficult a specific speaker was to understand in general terms (27%), and in about 30% of the total comprehensibility comments focused on grammar (e.g., "complex structures using the conditional", "no grammar errors to distract") and vocabulary (e.g., "not enough vocabulary", "use of 'valise' [instead of 'suitcase'] influenced my rating"). These results were confirmed by the analyses of the factors that the teachers identified as being important in their scoring decisions at the end of the testing session. For accent, all three teachers listed "pronunciation of vowels and consonants", "word stress", and "natural sounding rhythm" as the factors most affecting their judgments. For comprehensibility, all

unanimously identified "lexical errors" and "grammatical errors" as important.

To summarize, regression analyses carried out using 60 novice raters' judgments revealed that accentedness ratings were best explained using the dimensions of word stress and rhythm while comprehensibility ratings were best accounted for by the dimensions of word stress, type frequency, and grammatical accuracy. Analyses of the experienced teachers' reports yielded complementary findings. Namely, 54% of all rater comments for accentedness focused on the three most frequently coded categories of segmental errors (vowels and consonants), syllable errors, and sounding nativelike. In contrast, 57% of their comments for comprehensibility encompassed the three coded categories of ease of understanding, grammar, and vocabulary.

Discussion

The results of this study offer direct evidence that accent and comprehensibility (one measure of understanding) are overlapping yet distinct constructs. These constructs are complex and are associated with many linguistic factors drawn from the domains of phonology, fluency, lexis, grammar, and discourse. Fine-grained aspects of segmental accuracy, including errors of syllable structure, appear to factor into listeners' accent judgments, but have less bearing on their perceived comprehension difficulties. In contrast, grammatical errors and some aspects of vocabulary in L2 learners' speech appear to distract listeners from attending to the message, but have less relevance to their perception of L2 accent. This distinction between accent and comprehensibility was obtained both in quantitative analyses of the scores assigned by novice raters and in smaller-scale qualitative analyses of experienced raters' comments about the factors that they heeded when assigning scores. This distinction also emerged in experienced raters' indications of the linguistic factors most influencing their scoring decisions at the end of the session: "pronunciation of vowels and consonants" and "naturally sounding rhythm" unanimously featured in relation to accent, whereas "grammatical and lexical errors" were most relevant to comprehensibility. In addition, "storytelling elements and cohesion" and features of listener background characteristics that can facilitate understanding L2 speech (e.g., familiarity with the speakers' L1) were only mentioned in reference to comprehensibility.

It is not surprising that several factors from the domain of phonology appear to contribute to listeners' perception of a foreign accent. This is consistent with previous research examining phonological influences on accent (e.g., Anderson-Hsieh et al., 1992; Kang, 2010; Kang, Rubin & Pickering, 2010; Munro & Derwing, 1999). This finding is also in line with research on L2 teachers, who in teaching pronunciation place particular emphasis on segmental aspects of speech (Foote, Holtby & Derwing, 2011), and on L2 learners, who associate accent with various aspects of L2 phonology (Derwing, 2003). Finally, this result confirms the intuitive judgment of many listeners, who are remarkably sensitive to even slight deviations of speech from their local variety (Munro et al., 2010), that accent is linked to particular ways in which individual sounds, syllables, words, and longer elements of discourse are produced. It is also not surprising that for novice and experienced raters, different aspects of L2 phonology were related to the perception of accent. For novice raters, word stress and rhythm likely seemed most perceptually salient, given the wide cross-language difference between English and French stress placement and syllable reduction (discussed below). In contrast, for experienced raters (who were not simply rating speech but were asked to verbalize some of the reasons for their scoring decisions), segment and syllable structure errors appeared most serious. This is consistent with a typical instructional emphasis on segmentals in L2 classrooms and teachers' general lack of knowledge about or confidence in teaching prosody (Derwing, 2003; Foote et al., 2011). This finding may also reflect the fact that prosodic errors, compared to segmental errors, are difficult to describe without reference to specialized terminology and that examples of noticeable prosodic features are difficult to represent orthographically in written comments.

Unlike accent, comprehensibility appeared to be mostly, although not exclusively, linked to aspects of grammar and vocabulary in L2 speech. With respect to grammar, this finding extends the previous research citing the relationship between grammatical accuracy and listener evaluations of L2 speech, namely, that ungrammatical sentences negatively affect comprehensibility (Fayer & Krasinski, 1987; Varonis & Gass, 1982), that grammatical errors are associated with comprehensibility (Munro & Derwing, 1999), and that listeners find grammatical errors in L2 speech annoying and serious (Derwing, Rossiter & Ehrensberger-Dow, 2002). Taken together with these previous results, the grammar-comprehensibility link shown here suggests that listeners are distracted by grammatical errors from attending to the message in L2 speech, which makes comprehension more effortful. With respect to vocabulary, the current findings show that richer, more varied lexical content of L2 speech (i.e., greater type frequency, or a larger number of unique content words) is associated with higher comprehensibility ratings. This finding complements earlier demonstrations that L2 speakers' familiarity with L2 vocabulary in a speaking task can impact the quality of their L2 productions and, in turn, the severity of listeners' assessments (Munro & Derwing, 1994) and that listeners rely on semantic

context, including the lexical content of utterances, to assign speech ratings (Gass & Varonis, 1984). However, interpretations of the vocabulary-comprehensibility link should be cautious at best because type frequency, a measure of lexical richness, was also strongly associated with MLR, a measure of fluency ($r = .88$), and with story breadth, a measure of discourse complexity ($r = .75$). This suggests that richer L2 vocabulary is also linked to more fluent word retrieval and articulation and to more complex discourse structure, and that listeners may consider all these features in judging L2 comprehensibility.

Word stress was the dimension that emerged as common to both accent and comprehensibility in the analyses of novice rater judgments. The pervasive influence of word stress on both accent and comprehensibility in this study is to be expected, given that word stress is non-contrastive in French (Vaissière, 1991) and that unlike English stress-timed rhythm, which is characterized by regular alternations between stressed and unstressed syllables, French syllable-timed rhythm generally lacks such alternations (Ramus, Nespor & Mehler, 1999). Indeed, native French speakers often show "stress deafness" when asked to distinguish linguistic stimuli that solely differ in stress (e.g., Dupoux, Peperkamp & Sebastián-Gallés, 2001). Therefore, "unpredictable" stress placement (word stress) and alternations in stress (rhythm) in English would pose a major learning challenge for L1 French speakers, regardless of proficiency level, and would be a salient factor influencing listeners' perception of both accent and comprehensibility. And judging from the sheer number of learners from different L1 backgrounds for whom stress (and rhythm) generally pose a problem (e.g., Italian, Tagalog), it could be a much more global feature distinguishing between different L2 accent and comprehensibility levels. In fact, recent research shows that aspects of word stress and rhythm (with other prosodic features) account for up to 50% of the variance in accent judgments for L2 speakers from varied L1 backgrounds (Kang, 2010; Kang et al., 2010).[4]

Conclusions

Renewed interest in L2 pronunciation and speaking has shifted the instructional focus away from acquiring a nativelike accent to the more realistic goal of being communicatively effective (Derwing & Munro, 2009). However, assessment practice for bilinguals and multilinguals in both academic contexts (e.g., Levis, 2006) and workplace settings (e.g., Lacey, 2011) has yet to follow suit inasmuch as the majority of current L2 oral proficiency scales conflate the "conflicting" dimensions of accent and ease of understanding in their descriptors, particularly at the higher end of the scales (Isaacs & Trofimovich, in press). The Cambridge ESOL Common Scale for Speaking, for example, explicitly links the presence of a perceptually salient accent to pronunciation that is "unintelligible" or difficult to understand (University of Cambridge ESOL Examinations, 2008). The Common European Framework of Reference Scale of Phonological Control combines descriptions of easily understandable speech and a noticeable foreign accent in the same band descriptor (Council of Europe, 2001). The results cited here provide direct evidence that could help test developers, raters, researchers, teachers, and members of the general public to disentangle accent from different aspects of communicative effectiveness, including comprehensibility. What remains to be done is to validate the current findings with other measures of L2 speech and with other groups of L2 speakers, including multilinguals, in order to determine which influences on accent and comprehensibility are specific to a given sample of participants, specific measures, and a particular task and which ones cut across contextual variation. The eventual goals of such research would be to inform teachers about the most effective instructional targets in L2 pronunciation instruction, to describe communicative effectiveness with greater precision in rating scales, and ultimately to clarify both social and cognitive consequences of being a non-native speaker.

[4] A reviewer pointed out that an influential contribution of stress to both accent and comprehensibility could also be related to the fact that stress is one of the most structural and hierarchical aspects of phonology (e.g., in metrical phonology). This interesting hypothesis needs to be explored in future research.

References

Albrechtsen, D., Henriksen, B., & Færch, C. (1980). Native speaker reactions to learners' spoken interlanguage. *Language Learning, 30,* 365–396.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42,* 529–555.

Boersma, P., & Weenink, D. (2010). *Praat: Doing phonetics by computer* [computer program] (version 5.1.29). www.praat.org.

Cartier, F. A. (1975). Criterion-referenced testing in language skills. In L. A. Palmer & B. Spolsky (eds.), *Papers on language testing 1967–1974*, pp. 19–24. Washington, DC: TOEFL.

Clopper, C. G., & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics, 32,* 111–140.

Cobb, T. (2000). The compleat lexical tutor [website]. http://www.lextutor.ca.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Creswell, J. W., & Plano-Clark, V. (2011). *Designing and conducting mixed-methods research* (2nd edn.). Thousand Oaks, CA: Sage.

Derwing, T. M. (2003). What do ESL students say about their accents? *The Canadian Modern Language Review, 59,* 547–566.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42,* 476–490.

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics, 29,* 359–380.

Derwing, T. M., Rossiter, M. J., & Ehrensberger-Dow, M. (2002). "They speaked and wrote real good": Judgements of non-native and native grammar. *Language Awareness, 11,* 84–99.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54,* 665–679.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness". *Journal of the Acoustical Society of America, 110,* 1606–1618.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37,* 313–326.

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal, 29,* 1–22.

Gass, S., & Varonis, E. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 34,* 65–89.

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38,* 201–223.

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English speaking graduate students. *Canadian Modern Language Review, 64,* 555–580.

Isaacs, T., & Thomson, R. I. (in press). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*.

Isaacs, T., & Trofimovich, P. (2010). Falling on sensitive ears? The influence of musical ability on extreme raters' judgments of L2 pronunciation. *TESOL Quarterly, 44,* 375–386.

Isaacs, T., & Trofimovich, P. (in press). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition, 34 (3)*. doi:10.1017S0272263112000150.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System, 38,* 301–315.

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology, 28,* 441–456.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal, 94,* 554–566.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.

Lacey, M. (2011). In Arizona, complaints that an accent can hinder a teacher's career. *The New York Time*s (September 25, 2011), p. A18.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46,* 1093–1096.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39,* 369–377.

Levis, J. M. (2006). Pronunciation and the assessment of spoken language. In R. Hughes (ed.), *Spoken English, TESOL and applied linguistics: Challenges for theory and practice*, pp. 245–270. New York: Palgrave Macmillan.

Lippi-Green, R. (2011). *English with an accent: Language, ideology and discrimination in the United States* (2nd edn.). London: Routledge.

Mallows, C. L. (1964). Some comments on C_p. *Technometrics, 15,* 661–675.

Martin, J. R., & Rose, D. (2003). *Working with discourse: Meaning beyond the clause*. London: Continuum.

Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal, 20,* 38–51.

Munro, M. J., & Derwing, T. M. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing, 11,* 253–266.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49,* 285–310.

Munro, M. J., Derwing, T. M., Burgess, C. S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication, 52,* 626–637.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences, 12,* 237–241.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73,* 265–292.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgements of truth. *Consciousness and Cognition, 8,* 338–342.

Robinson, G. C., & Stockman, I. J. (2009). Cross-dialectal perceptual experienced of speech-language pathologists in predominantly Caucasian American school districts. *Language, Speech, and Hearing Services in Schools, 40,* 138–149.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of non-native English-speaking teaching assistants. *Research in Higher Education, 33,* 511–531.

SAS Institute Inc. (2004). *SAS/STAT® 9.1 user's guide*. Cary, NC: SAS Institute Inc.

Schairer, K. E. (1992). Native speaker reaction to non-native speech. *Modern Language Journal, 76,* 309–319.

Southwood, M. H., & Flege, J. E. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics, 13,* 335–349.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (ed.), *New directions in discursive processing*, pp. 53–120. Norwood, NJ: Ablex.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage.

Trofimovich, P., Gatbonton, E., & Segalowitz, N. (2007). A dynamic look at L2 phonological learning: Seeking processing explanations for implicational phenomena. *Studies in Second Language Acquisition, 29,* 407–448.

University of Cambridge ESOL Examinations (2008). *Certificate of Proficiency in English: Handbook for teachers*. Cambridge: UCLES.

Vaissière, J. (1991). Rhythm, accentuation and final lengthening in French. In J. Sundberg, L. Nord & R. Carlson (eds.), *Music, language, speech and brain* (Wenner–Gren International Symposium Series 59), pp. 108–120. London: Macmillan.

Varonis, E. M., & Gass, S. (1982). The comprehensibility of non-native speech. *Studies in Second Language Acquisition, 4,* 114–136.

Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. Oxford: Oxford University Press.

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1235–1253.