Chapter Title: The Role of Pronunciation in the Assessment of Second Language Listening Ability

Chapter Author(s): Elvis Wagner and Paul D. Toth

Book Title: Second Language Pronunciation Assessment

Book Subtitle: Interdisciplinary Perspectives

Book Editor(s): Talia Isaacs and Pavel Trofimovich

Published by: Multilingual Matters / Channel View Publications

Stable URL: https://www.jstor.org/stable/10.21832/j.ctt1xp3wcc.9

5   The Role of Pronunciation in the Assessment of Second Language Listening Ability

Elvis Wagner and Paul D. Toth

Introduction

Traditionally, second language (L2) pronunciation has been operationalized as a component of speaking ability, and efforts to assess L2 learners' pronunciation have focused on the test taker's spoken production. But pronunciation also plays a role in the assessment of L2 listening ability. L2 listening tests almost invariably utilize recorded spoken texts to assess the test taker's comprehension. Isaacs (2014) argues that it is essential in L2 speaking assessment to define and clarify the role of pronunciation within this construct. However, we believe that it is also necessary to do so in the assessment of L2 listening, because how the speakers of L2 texts articulate their utterances can impact on comprehensibility and, consequently, test taker performance. This chapter explores a number of points related to this issue, including: how the spoken texts used in L2 listening tests are chosen; the effects of scripted versus unscripted texts; the organization, phonology and fluency characteristics of spoken texts; and how these issues impact on construct validity. The chapter then presents an empirical study investigating L2 test takers' beliefs about the nature of spoken texts used in an L2 listening test.

Review of the Literature

Spoken texts used in L2 tests

In theory, the target language use (TLU) domain of interest dictates the types of spoken texts that will be used in L2 listening assessments (Bachman & Palmer, 1996). In other words, the goal of a test is to assess a test taker's language ability beyond the test context, and so if an L2 listening test

purports to assess listeners' ability to understand interactive, conversational spoken language (the TLU domain), then it should include spoken texts that involve interactive, conversational spoken language. The spoken texts used in the test should have similar phonological, linguistic, organizational, pragmatic and lexico-grammatical characteristics to real-world conversational language, and the pronunciation patterns in the test tasks should be similar to the pronunciation patterns of speakers in the TLU domain (Buck, 2001; Wagner, 2013b, 2014; Wagner & Toth, 2014). Using spoken texts in the test that have the same characteristics as spoken texts in the TLU domain should result in more valid inferences about the test takers' ability in that domain, while using spoken texts that utilize formal, over-enunciated spoken language can result in a narrow operationalization of the construct, which can present threats to the validity of the test (Messick, 1989, 1996).

In practice, however, many of the spoken texts used in L2 listening tests are very different from the spoken texts that test takers will encounter outside the test-taking context. In other words, the spoken texts used in the tests are not representative of the texts in the TLU domain. Wagner (2013b, 2014) describes how texts that are used in L2 listening tests are developed. Generally, test developers have test task specifications that dictate the number of texts to be used, their designated genre and length, and the number and types of response items to include for each. As a result, it is usually more efficient and practical for test developers to create a spoken text that corresponds to these specifications than to identify or record authentic texts. Thus, test developers usually create a scripted text (planned, written, revised, edited and polished) that is then read aloud by voice actors. The resulting scripted texts are often different from the unscripted, real-world spoken language of the TLU domain of interest.

Differences between scripted and unscripted texts

There has been extensive research on the characteristics of scripted and unscripted texts, and the results are summarized here (see Gilmore, 2007, for an analysis of how 'authentic' spoken texts differ from spoken texts commonly used in L2 materials). Wagner (2014), Wagner and Toth (2014) and Wagner and Wagner (2016) outline major categories of difference between texts that are planned, scripted and read aloud, and unplanned, spontaneous speech, where the speaker composes and utters the text virtually simultaneously. These include: articulatory/phonological characteristics (i.e. connected speech); organizational/planning characteristics; spoken grammar; oral lexicon; rate of speech; and hesitation phenomena. Among these, connected speech is the category most relevant for investigating pronunciation's role in L2 listening assessment, but the other five categories will also be briefly reviewed here as they are also secondarily related to the issue of pronunciation in L2 listening assessment.

Connected speech

The process of articulating rapid speech leads to phonological modifications that differ from citation forms in oral language. Brown and Kondo-Brown (2006) explain that articulated speech is affected by a number of processes, including word stress, sentence timing and stress, reduction, citation and weak forms of words, elision, intrusion, assimilation, juncture and contraction. These processes result in connected speech, which is typical of most real-world speaking events. It is also commonly accepted that the formality of a speech event affects pronunciation and connected speech: the more formal the event, the more careful and conscious the speaker is of enunciation, so that connected speech is less likely to occur. However, Brown and Kondo-Brown (2006) stress that, although connected speech is more common in informal contexts, it occurs in all registers and styles. Speakers can consciously attend to their pronunciation to reduce connected speech with a goal of clearer enunciation and greater intelligibility (Ito, 2006; Ladefoged & Johnson, 2010; Mora & Darcy, this volume). This is especially true when the speaker reads a text aloud, rather than composing and speaking a text simultaneously (Chafe, 1982; Haviland & Clark, 1974).

Organizational/planning characteristics

When scripted texts are read aloud, the speaker does not have to plan and speak simultaneously. Consequently, these texts often have a more formal, linear organization in the presentation of propositional content which reflects the planning and editing of the writer(s) of the text. In contrast, unscripted spoken texts tend to be less linearly organized because of the cognitive constraints involved in composing and uttering the text simultaneously (Chafe, 1982). As a result, unscripted spoken texts usually have many more digressions, false starts, redundancies and hesitation phenomena, including filled and unfilled pauses (Chafe, 1982; Haviland & Clark, 1974; Rost, 2011). This idea of hesitation phenomena is examined in more detail below.

Spoken grammar

Numerous researchers (e.g. Chafe, 1982, 1985; Halliday, 1985) have described how the grammatical characteristics of written (scripted) language differ from unscripted spoken language, including a greater use of complex syntactic structures like embedded clauses, agentless passives and nominalizations. In contrast, unscripted spoken texts generally have shorter idea units, more run-on sentences and less complex syntax. Indeed, corpus linguistic research has documented marked differences between the grammatical system of real-world spoken language and formal, scripted (written) language (Biber, 1988, 2006; Biber & Gray, 2013; McCarthy & Carter, 1995, 2001).

Oral lexicon

Similarly, Brown (1995) and Chafe (1985) have explained how written language generally contains less slang and colloquialisms than more

spontaneous, spoken language, in part because oral language tends to be less formal than written language. In addition, corpus linguists have shown that a speaker's oral and written lexicons can differ markedly from one another. McCarthy (2010) found that there was only about a 65% overlap between the 2000 most common words in a spoken versus a written corpus, and noted the importance of 'turn-openers' and 'turn-closers', like *yeah, oh,* and *mm,* in informal conversation. While these turn-openers and turn-closers can perform a number of functions, including as backchannels, and as turn-holding interactional strategies that the speaker uses while considering what to say and how to say it, they are much less common in scripted language, and most L2 learners (and teachers) would probably identify them as slang or colloquialisms.

Speech rate

Speech rate is generally defined as a measure of how quickly a person is speaking, and is often measured as the number of words, syllables or phonemes divided by the duration of the speech (Cucchiarini *et al.*, 2010). It is widely accepted that the rate of speech can affect L2 listening comprehension; the research has found that spoken texts delivered at a faster speech rate are more difficult for L2 listeners to comprehend than texts delivered at a slower rate (e.g. Griffiths, 1992; Kelch, 1985; Zhao, 1997). This is generally attributed to increased processing time, in that a slower speech rate allows the L2 listener more processing time. However, another consideration relevant to this study is the idea that a speaker's rate of speech can be influenced by his or her attempt to enunciate carefully. The fact that it is easier for a speaker to enunciate clearly when he or she speaks more slowly might make the text more intelligible (and comprehensible). This could partly explain why oral texts produced by highly proficient speakers of the target language and delivered at a faster rate are more difficult for L2 listeners to comprehend than texts delivered at a slower rate.

Hesitation phenomena

Hesitation phenomena are the filled and unfilled pauses, false starts, hesitations, redundancies and repeats that are characteristic of spontaneous spoken language. Because of the real-time nature of unplanned speech, where the speaker composes and utters speech almost simultaneously, hesitation phenomena can occur as the speaker searches for what to say and how to say it (Chafe, 1985; Wagner, 2014), or decides to rephrase an utterance (McCarthy, 2005). In spontaneous speech, filled pauses seem to be more common than unfilled pauses, while in scripted texts that are read aloud, unfilled pauses seem to be more common than filled pauses (Cucchiarini *et al.*, 2010; Wagner & Wagner, 2016).

How these hesitation phenomena affect L2 listeners' ability to understand spoken texts is a matter of debate. On the one hand, Griffiths (1991) explains that L2 learners might have difficulty in processing some types of hesitation

phenomena as they try to assign semantic meaning to filled pauses (*uh, you know*). Indeed, empirical studies such as Voss (1979) and Griffiths (1991) have found that L2 learners are less able to comprehend spoken texts with filled pauses than texts without them. Freedle and Kostin (1999) also examined the influence of hesitation phenomena on L2 listening performance, and found that texts with both filled pauses (e.g. *um* or *er*) and unfilled pauses of one second or more were actually more difficult for L2 test takers to comprehend than texts without such pauses. This was contrary to what they had hypothesized, and they concluded that 'apparently any disruption in the coherent reception of a speaker's ideas made it harder to process the message' (Freedle & Kostin, 1999: 18). However, while the findings of Voss (1979) and Griffiths (1991) focused on filled pauses, Freedle and Kostin grouped filled and unfilled pauses together because of their low frequency in the data, so the effects of a particular type of pause could not be ascertained. In addition, their pauses were purposefully inserted by the trained native English speakers of the texts to make them sound more authentic, so it is unclear whether these pauses were truly similar to and representative of the types of pauses found in real-world spoken language.

Alternatively, one could argue that hesitation phenomena might actually facilitate comprehension, in that the pauses and false starts would allow L2 listeners 'extra time to process what they hear' (Vandergrift & Goh, 2012: 154). This would possibly explain the findings reported earlier that a slower speech rate often leads to increased L2 comprehension. Indeed, Blau (1990) found that L2 learners scored higher on listening tests involving spoken texts with blank pauses mechanically inserted at normal discourse boundaries, than did learners hearing the same texts without these pauses. Similarly, Parker and Chaudron (1987) found that texts that included repetitions and redundancies (i.e. repeated phrases and clauses within the text) led to increased L2 listening comprehension. Again, this might be because these features effectively slowed down the speech rate. However, if the L2 listener is actively trying to decode the filled pauses and extract semantic information from them (as Voss, 1979, and Griffiths, 1991, found), or if the hesitation phenomena disrupt text processing, then the resulting slower speech rate might not actually benefit L2 listeners.

In summary, there is a broad consensus that the phonological/articulatory characteristics of unplanned spoken texts (i.e. connected speech that includes reduction, elision, intrusion, assimilation, juncture, etc.) make unscripted spoken texts more difficult for L2 listeners than the over-enunciated speech typical of scripted texts. Similarly, evidence suggests that spoken texts delivered at a faster rate are generally more difficult for L2 listeners than spoken texts delivered at a slower rate. However, even this issue is far from clear, because the hesitation phenomena typical of unplanned spoken texts (i.e. pauses and redundancies) can also serve to reduce the speech rate and, thus, potentially facilitate comprehension. The literature also suggests

that the organizational patterns, spoken grammar and oral lexicon characteristic of unscripted, spontaneous language can present difficulties for L2 listeners. Finally, spoken texts with the properties of unplanned spoken language might be more difficult for L2 listeners if they have not been exposed to or taught this kind of language. Indeed, the literature suggests that it is not necessarily the characteristics of unscripted spoken language that present difficulties for L2 listeners, but rather a lack of exposure to such texts, or instruction on and strategies for how to process them. Numerous researchers (e.g. Dupuy, 1999; Field, 2008; Gilmore, 2007; Meinardi, 2009; Wagner, 2014; Wagner & Toth, 2014) have argued that exposing L2 learners to unscripted oral texts and drawing learners' attention to their features so that learners will notice them in subsequent input is effective in developing the ability to comprehend unscripted, real-world spoken language. Yet for many L2 learners, especially foreign language learners, much of the spoken input they are exposed to comes from L2 textbook materials, and the types of spoken texts used in L2 textbooks have been found to differ extensively from real-world spoken language (Flowerdew & Miller, 1997; Gilmore, 2004, 2007; Thompson, 2003). Similarly, the spoken texts used in L2 listening tests seem to consist almost entirely of scripted texts that are read aloud (Wagner, 2013b), even though it is entirely feasible to use unscripted or semi-scripted spoken texts (Clark, 2014).

Learners' attitudes towards the use of authentic spoken texts

Field (2008) and Meinardi (2009) have argued that using authentic, unscripted texts for L2 listening instruction can enhance positive affect and motivation, which in theory should lead to more positive learning outcomes. This issue has received increasing attention in the field, and materials developers have seemed more concerned about the authenticity of their listening tasks in recent years. Gilmore (2004) examined the spoken texts used in ESL/EFL textbooks, and while most used inauthentic texts with few characteristics of real-world spoken language, more recent textbooks seem to be incorporating at least some natural discourse features in their texts. The issue has also begun receiving more attention in ESL/EFL teacher training materials (e.g. Brown, 2012). While many researchers have argued for authentic, unplanned spoken texts in L2 classrooms, some have expressed concern about the level of difficulty they entail, especially for lower proficiency learners. Guariento and Morley (2001) asserted that using such texts can lead to confused and frustrated learners and ultimately poor learning outcomes. Similarly, Richards (2006) cautioned about the 'myth of authenticity', arguing that authentic spoken texts for L2 instruction were difficult to obtain and even more difficult to implement without substantial modification.

A surprisingly limited number of empirical studies investigating language learners' attitudes towards unscripted texts in L2 materials have been

conducted, and the results are mixed, in part because the studies have used very different methodologies and examined learners of differing proficiency levels, but also perhaps because they have had relatively small participant numbers. Kmiecik and Barkhuizen (2006) surveyed 17 ESL learners, and found more negative attitudes towards authentic texts, because learners struggled with the speed of the input and the difficulty of the vocabulary. Peacock (1997) surveyed 31 EFL learners, who reported that even though they found authentic listening materials more motivating than artificial materials, they also found them less interesting. Gallien *et al.* (2000) surveyed 48 learners of French and German as a foreign language, who reported that simplified texts were more 'appealing' than authentic spoken texts.

Furthermore, one justification commonly made for authentic spoken L2 materials is that such material will be motivating for students. However, this is not fully supported by empirical data, and there seem to be no studies that investigate learners' attitudes towards authentic, unscripted spoken input in L2 listening tests. Perhaps more importantly, there is also very little empirical evidence examining the extent to which L2 learners are even aware of the differences between unscripted, real-world spoken language, and the scripted and polished spoken texts often found in L2 materials. Nonetheless, test takers' attitudes and beliefs about testing materials can have a real influence on their scores, as test taker affect can impact motivation and performance. In addition, the materials used in L2 tests (and the test takers' attitudes towards them) can contribute to a washback effect, both positively and negatively, on stakeholders including test takers, teachers and educational systems (Buck, 2001; Wagner, 2014; Wagner & Wagner, 2016). If a high-stakes test uses unscripted spoken texts for L2 listening assessment, then it is more likely that curriculum planners, materials developers and classroom teachers will likewise use unscripted spoken texts in their materials for L2 learners.

The Current Study

The current study explores test takers' awareness and beliefs about the types of spoken texts used in an L2 listening test. It is part of a larger investigation of unscripted spoken texts in L2 listening comprehension assessment. As reported in Wagner and Toth (2014), two comparable groups of L2 Spanish learners took an L2 listening test. For one group, the spoken texts were unscripted, and consequently had many of the organizational, phonological and fluency characteristics found in spontaneous, real-world language, as well as extensive instances of connected speech and hesitation phenomena. The second group took the same L2 listening test, except that the spoken texts were scripted and lacked most of the characteristics of unplanned spoken language. As hypothesized, the group of 86 test takers that listened to the scripted texts scored 8.4% higher on the listening

comprehension test than the group of 85 test takers who listened to the unscripted texts, and this difference was statistically significant (Wagner & Toth, 2014). However, we were not only interested in how the two different groups would perform on the test, but we also wanted to examine the extent to which the test takers were aware of the organization, phonology and fluency characteristics of the different spoken texts. Thus, after completing the test, test takers in both groups were surveyed through the use of a written questionnaire about the spoken texts used in the test. The following research question was addressed: What are the test takers' beliefs about the characteristics of the spoken language used on an L2 listening comprehension test?

Methodology

For the questionnaire data, the independent variable was the type of audio-text used in the listening comprehension test: 'unscripted' versus 'scripted'. The dependent variables were group scores on five different sub-scales of the questionnaire that asked the participants about their views of the spoken texts used in the listening test. A series of independent sample t-tests assessed how the independent variable affected the group scores on the sub-scales of the questionnaire.

Participants

This study involved 171 learners of Spanish as a foreign language (SFL) at a large American public university. All were students in an intermediate-level, fourth-semester Spanish course entitled 'Conversational Review', which focused on speaking and listening skills. There were 14 classes of 'Conversational Review' in the study, which were randomly assigned to one of two groups: seven classes were assigned to the unscripted group, and seven to the scripted group. Of the 85 test takers in the unscripted group, 81 listed English as their L1, and four listed a language other than English. None had Spanish as their L1, although two listed a Romance language (Romanian). For the unscripted group, the average age was 20.24 years and 59% were female. For the 86 participants in the scripted group, 78 listed English as their L1, and seven listed a language other than English. None had Spanish or a Romance language as their L1. The average age of the group was 20.45 years and 70% were female. A self-assessment was used to examine if the two groups had comparable perceptions of their L2 Spanish proficiency. The test takers used a six-point scale (1 = lower beginner, 2 = upper beginner, 3 = lower intermediate, 4 = upper intermediate, 5 = lower advanced, 6 = upper advanced). The two groups' self-assessments were very similar: 3.49/6.00 for the unscripted group and 3.53/6.00 for the scripted group.

Spoken texts

Two spoken texts were created for the study described in Wagner and Toth (2014). Two female L1 speakers of Peruvian Spanish were used; they were given the basic outlines for performing a role-play to create the two texts: one called 'A Room for Rent', and another called 'A Friend Goes on Vacation'. For the former, one speaker was a university student seeking to rent a room from the other; for the latter, one speaker gave instructions to the other for taking care of her house while she went on vacation. After reading the role-play instructions and considering what they might say for a few moments, the two speakers were instructed to speak as naturally as possible for approximately three to four minutes. The speakers then recorded the two texts.

After the unscripted texts were completed, the researchers transcribed them, and then revised and edited the transcripts to remove the pauses, false starts, hesitations, redundancies, overlaps and backchannels. This resulted in fewer speaker turns in the texts, and a more linear organizational scheme. Using these edited and polished transcripts, the same two native Spanish speakers were then instructed to read the transcripts aloud, and to simulate the types of spoken texts found in L2 classroom materials. They were instructed to be conscious of enunciating clearly, to avoid connected speech and overlapping with the other speaker, and not to speak too rapidly.

The resulting two versions of the spoken texts were thus equivalent in topic, content and information, and were spoken by the same speakers. They differed, however, in the presence or absence of connected speech and hesitation phenomena and their related organizational characteristics, as virtually all instances of overlapping talk, filled pauses, repeated phrases, backchannels and exclamatives from the unscripted texts were absent in the scripted versions (see Wagner & Toth, 2014). Thus, in the scripted text, related propositions spread over two or more turns (with interruptions from the other speaker) were consolidated into single turns, and all false starts, repetitions, backchannels, exclamatives and filled pauses were simply deleted. It should be noted, however, that the vocabulary and propositional content in the two texts was nearly identical. Furthermore, there was no slang or colloquial language used in either version of the texts, apart from the fillers and backchannels in the unscripted text.

Instruments

After the spoken texts were created, eight multiple-choice listening comprehension items were developed for each of the texts, resulting in a 16-item test. A 21-item questionnaire was administered to examine the test takers' beliefs and opinions about the spoken input they heard on the exam. It was developed and validated based on Wagner's (2010, 2013a) suggestions for

applied linguistics survey research. The questionnaire used Likert items with five choices: 5 = 'strongly agree', 4 = 'agree', 3 = 'no opinion', 2 = 'disagree' and 1 = 'strongly disagree'. The five sub-scales of the questionnaire (described below) were based on a review of literature on the use of unscripted spoken texts in L2 teaching and testing (e.g. Gilmore, 2007; Wagner, 2013b).

The first sub-scale, 'Authentic versus Modified', was a five-item, holistic measure of beliefs about whether the recordings used in the test were real-world spoken texts or scripted texts created specifically for L2 learners. For example, one item asked: 'The texts required me to listen to authentic spoken language, the same type of spoken language that is found in real life.' A group mean above 3 on this sub-scale indicates that test takers thought the spoken texts were authentic and unscripted, while a group mean below 3 indicates that they thought the texts were scripted and modified, and created for L2 learners.

The other four sub-scales, each composed of four items, asked about specific characteristics of the text. The second sub-scale, 'Pronunciation', asked test takers about how the speakers enunciated their speech. For example, one item asked: 'It was hard to understand the speakers because they did **not** enunciate well and did **not** speak clearly.' (This item, and a number of other negatively worded items were reverse-coded in the analysis.) Other options in this sub-scale included: 'The speakers spoke clearly and used very clear pronunciation, which made it easier to understand them'; 'The speakers pronounced each word clearly and distinctly'; and 'The speakers' pronunciation in the texts was similar to native Spanish speakers' pronunciation in real-life conversations.' A group mean above 3 on this sub-scale indicates that test takers thought the speakers enunciated normally and that their pronunciation was similar to real-world spoken language, while a group mean below 3 indicates that they thought the speakers over-enunciated and spoke extra clearly so that L2 listeners could understand them.

The third sub-scale, 'Speech Rate', asked about how quickly the speakers spoke in the texts. For example, one item asked: 'The speakers on the spoken texts spoke quickly, the same rate that native speakers normally use with each other.' A group mean above 3 on this sub-scale indicates that the test takers thought the speakers spoke quickly, similarly to highly proficient speakers in real-world contexts, while a group mean below 3 indicates that they thought the speakers spoke artificially slowly and deliberately.

The fourth sub-scale, 'Pauses and False Starts', asked test takers about the extent to which the speakers in the texts had hesitation phenomena in their speech. For example, one item asked: 'The speakers often had a lot of pauses, fillers (things like *'eh …', 'em …', 'este …', 'tú sabes …'*), and false starts in their speech.' A group mean above 3 on this sub-scale indicates that the test takers thought the spoken texts had pauses, fillers and false starts like those in real-life unplanned spoken communication, while a group mean below 3 indicates that they thought the spoken texts were rehearsed and read aloud and did not include hesitation phenomena found in unscripted language.

The fifth sub-scale, 'Use of Slang', asked whether test takers thought the speakers used slang in their speech. For example, one item asked: 'The speakers used slang and informal expressions that are found in real-life language.' As stated above, the vocabulary used in the two texts was virtually identical, and thus there was no difference in the amount of slang or colloquial language used in the scripts. Nevertheless, we decided to include this sub-scale on the questionnaire in order to examine the extent to which the learners associated unscripted language with colloquial or non-standard speech. This seemed relevant, given our experience with language learners referring to any non-standard speech as *slang*, with a somewhat negative connotation. A group mean above 3 on this sub-scale indicates that the test takers thought the speakers did use slang and colloquial language, while a group mean below 3 indicates that they thought the speakers did not use slang.

The initial 21-item questionnaire was created, and then piloted with a group of 14 learners in a 'Conversational Review' class. After completing the questionnaire, the 14 test takers were surveyed about the questionnaire items, and asked about any items they found particularly difficult or problematic. In addition, a statistical analysis of the responses was conducted. Based on these qualitative and quantitative analyses, a number of items were revised until the questionnaire resulted in its final form. A complete list of the questionnaire items is provided in the Appendix to this chapter.

Procedures

The researchers went to the 14 different 'Conversational Review' classes to administer the test and post-test questionnaire. Because these were low-proficiency learners, the directions for the listening comprehension test were given in English both on the audio-recording and in the test booklet. The test items were written in Spanish. The test took about 20 minutes to complete, after which the test takers completed the 21-item questionnaire. The questionnaire items were ordered randomly and were written in English. Test takers circled numbers corresponding with an answer of 'strongly disagree', 'disagree', 'no opinion/don't know', 'agree', and 'strongly agree'. The questionnaire took about 10 minutes.

Test takers were not told before they took the test or questionnaire what the purpose of the study was. Rather, they were told that the researchers were examining how L2 learners perform on a listening test. They did not know that there were two versions of the spoken texts used in the test, and thus when they completed the questionnaire they responded based only on the version that they had just heard.

Analyses

The internal consistency reliability for each of the five sub-scales on the questionnaire was estimated separately for each group using Cronbach's

alpha. The item-total correlation for each item with its overall sub-scale was also examined to see how reliably each item performed. Descriptive statistics were computed to examine the central tendency and dispersion of the two groups on each of the sub-scales. A series of independent sample *t*-tests was then conducted to see if the two groups' beliefs and impressions of the two texts differed; that is, the means for the two groups' scores on the five sub-scales (authentic versus modified, pronunciation, speech rate, pauses and false starts, and use of slang) were compared to see if the groups' beliefs about the five variables differed.

Results

Beliefs about the spoken texts

While 20 of the 21 items on the questionnaire performed well statistically, the item-total correlation for item 21 (part of the 'pronunciation' sub-scale) was very low for both groups. Test takers were asked: 'The speakers' pronunciation in the texts was similar to native Spanish speakers' pronunciation in real-life conversations.' In reviewing this item, it became apparent that it differed from the other pronunciation items in that it asked whether the speakers' pronunciation was similar to that of native speakers, while the other four items focused on clarity and enunciation. Because the native speaker item did not seem to be reliably measuring the same construct, it was deleted from the rest of the analysis.

For both the unscripted and scripted groups, each of the five sub-scales had a moderately high internal consistency. For the unscripted group, the reliability coefficient for each sub-scale was: authentic versus modified, $\alpha = 0.74$; pronunciation, $\alpha = 0.81$; speech rate, $\alpha = 0.80$; pauses and false starts, $\alpha = 0.68$; and use of slang, $\alpha = 0.80$. For the scripted group, the reliability for each sub-scale was: authentic versus modified, $\alpha = 0.80$; pronunciation, $\alpha = 0.73$; speech rate, $\alpha = 0.80$; pauses and false starts, $\alpha = 0.54$; and use of slang, $\alpha = 0.82$. While these reliability figures are relatively high, the coefficient for the pauses and false starts sub-scale for the scripted group is markedly lower.

The descriptive statistics for both groups' scores on the five sub-scales of the questionnaire were also calculated. As shown in Table 5.1, the mean scores on the five sub-scales are consistently higher for the unscripted group than the mean scores for the scripted input group. To reiterate, the means for each sub-scale are based on five-point scales with 3 as the mid-point, so a higher mean on these sub-scales indicates that the test takers thought the texts were more authentic, that the pronunciation was more like real life, that there was a more natural speech rate, that there were more pauses and false starts and, finally, that there was more slang.

Table 5.1 Descriptive statistics for the scripted and unscripted groups on the five sub-scales of the questionnaire ($n = 171$)

Variable	Authentic vs. modified		Pronunciation		Speech rate		Pauses and false starts		Use of slang	
	U	S	U	S	U	S	U	S	U	S
Mean rating	3.52	2.99	3.22	2.14	3.85	2.66	3.49	2.40	2.91	2.39
SD	0.55	0.67	0.86	0.63	0.69	0.81	0.64	0.49	0.62	0.60
Kurtosis	0.04	−0.81	−1.20	0.95	0.86	0.86	−0.49	0.62	−0.77	−0.72
Skewness	−0.42	−0.16	−0.26	0.91	−1.10	0.37	−0.11	−0.12	−0.13	−0.07
Reliability	0.74	0.80	0.81	0.73	0.80	0.80	0.68	0.54	0.80	0.82

Notes: U = unscripted group; S = scripted group. Mean ratings are based on a five-point scale. Reliability is calculated using Cronbach's alpha.

Between-group comparisons

In order to compare the two groups' means on the five sub-scales of the questionnaire, five independent-sample t-tests were conducted. Because using five t-tests raises the possibility of finding group differences when in fact there are none, a Bonferroni adjustment set the significant level for multiple comparisons at 0.01. The skewness and kurtosis figures given in Table 5.1 suggest that the data for both groups are normally distributed. Levene's test of homogeneity of variances was conducted on the five sub-scales and it was found that for only one of the sub-scales, the use of slang, could the variances be considered homogeneous. Therefore, on the other four sub-scales, the numbers reported in the t-tests will be for 'equal variances not assumed'. The two-tailed t-tests for all five comparisons of the two groups' scores on the questionnaire sub-scales were statistically significant: authentic versus modified, $t(168.64) = 5.70$, $p < 0.001$, $d = 0.86$; pronunciation, $t(154.04) = 10.45$, $p < 0.001$, $d = 1.43$; speech rate, $t(165.36) = 10.45$, $p < 0.001$, $d = 1.58$; pauses and false starts, $t(156.59) = 12.50$, $p < 0.001$, $d = 1.91$; and use of slang, $t(169) = 5.56$, $p < 0.001$, $d = 0.85$. As shown by Cohen's d-effect size values, the effect sizes for the five sub-scales were all large. These tests indicated that there was a statistically significant difference in the two groups' beliefs about the spoken input on the tests for each of the five sub-scales of the questionnaire.

Discussion

Our research question asked: 'What are the test takers' beliefs about the characteristics of the spoken language used on an L2 listening comprehension

test?' The results indicate that the two groups of test takers had very different beliefs about the texts based on whether they heard the unscripted or the scripted texts. Again, test takers were not informed about the purpose of the study before they took the test or questionnaire, and they did not know which type of text they had heard. Yet the two groups' responses to the questionnaire differed significantly on all five sub-scales.

The unscripted group's score was more than a half-point higher (3.52 versus 2.99) than the scripted group on the first sub-scale, which asked whether participants thought the texts they heard were authentic, natural and representative of real-world spoken language. This statistically significant result indicates that learners could indeed distinguish authentic spoken texts from those created especially for L2 learners.

The second sub-scale focused on the extent to which test takers thought the texts had the pronunciation patterns and characteristics found in real-life Spanish conversations (i.e. if the speakers enunciated on the texts as they would in real conversation). The unscripted group's score of 3.22 was more than a full point higher than the scripted group's score of 2.14, which was the lowest of any score on the five sub-scales. Thus, participants in the scripted group were well aware that the pronunciation they heard was different from real-world language, and that the speakers were enunciating more clearly than they would in a real-world context. Similarly, the unscripted group's score on the speech rate sub-scale was more than a full point higher than the scripted group (3.85 and 2.66, respectively), which yielded the largest difference for any of the five sub-scales. This meant that the unscripted group participants agreed with statements affirming that the speakers on the texts spoke quickly, as native speakers do when conversing. For the fourth sub-scale, which asked whether the spoken texts had pauses and false starts similar to those of real-world language, the unscripted group again scored more than a point higher (3.49) than the scripted group (2.40), suggesting that listeners perceived the hesitation phenomena that were present in the unscripted texts but virtually absent in the scripted texts. However, the reliability coefficient was much lower for the pauses and false starts sub-scale for the unscripted and scripted groups ($\alpha = 0.68$ and $\alpha = 0.54$, respectively), so the results must be interpreted with caution.

The results of the scores on the fifth and final sub-scale, 'use of slang', are difficult to interpret. This sub-scale asked participants about how much slang and colloquial language the speakers used in the text. The mean score of 2.91 was the lowest of the five sub-scales for the unscripted group and below, in fact, the mid-point of the sub-scale. Similarly, the mean score of 2.39 for the scripted group was the lowest of the five sub-scales. While this was the smallest difference in means for any of the five sub-scales, it was still statistically significant. These scores are difficult to explain because the vocabulary in the two texts was virtually identical, with no lexical

modifications made to the scripted text apart from the removal of filled pauses. One possible reason for the difference in scores is that, because the unscripted group perceived their text to be more natural, they might have assumed the speakers were using slang and colloquialisms (including the fillers such as *'um'*, *'este'* and *'o sea'*). Similarly, because the scripted group perceived the texts as being unnatural with overly formal enunciation, they might have assumed that the speakers would be less likely to use slang and colloquial language.

A limitation of this study is that while it compared two groups' beliefs about the texts they heard, each group heard only one type of text. A counterbalanced design in which each group heard and rated both types of texts would have been stronger. Nevertheless, our findings suggest that learners can distinguish spoken texts made especially for L2 learners from unplanned, unscripted speech that reflects real-world spoken language. There does not seem to be any literature that has specifically focused on L2 learners' ability to detect if a spoken text is scripted or unscripted, so these results must be seen as exploratory. As reported in Wagner and Toth (2014), the learners in the unscripted group scored lower on the comprehension test than the unscripted group. It is not surprising, then, that these learners would report that the texts seemed similar to the authentic spoken language of native speakers. Because of the learners' relative difficulty in comprehending and processing the text, we can speculate that they associated it with authentic, unscripted speech, and thus perceived the speakers as talking quickly, using slang and colloquialisms, not enunciating clearly, and employing numerous pauses and fillers. This would mirror the results of Kmiecik and Barkhuizen (2006), who found that ESL learners had more negative attitudes towards authentic spoken texts due to comprehension difficulties arising from a high speech rate and the use of unfamiliar vocabulary. Likewise, our results reflect Gallien *et al.*'s (2000) study, where FL learners found simplified texts more appealing than authentic texts, in part due to ease of comprehension.

It seems unlikely that the listeners in this study were conscious of many of the organization, phonology and fluency characteristics of the spoken texts while they were listening to them. Rather, at least for some participants, the items in the questionnaire likely forced them to think about the different characteristics, and test takers who had difficulty with the texts may have equated the challenge they faced with a particular text type. This would perhaps explain the anomaly of the 'use of slang' sub-scale. Again, the test takers in the unscripted group rated their texts as having more slang and colloquial language than test takers in the scripted group, even though the vocabulary was virtually identical in both text versions.

In order to examine this hypothesis, we carried out a post hoc analysis, in which we divided the two test taker groups into 'high-comprehenders', who scored above the median on the listening comprehension test, and

'low-comprehenders', who scored at or below the median. For the scripted group, there was no difference among high- and low-comprehenders' mean scores on any of the five sub-scales. For the unscripted group, however, the low-comprehenders' mean score (3.02, $SD = 0.56$) on the 'use of slang' sub-scale was significantly higher than the high-comprehenders' mean score (2.71, $SD = 0.71$), $t(83) = 2.31$, $p = 0.024$. In addition, the low-comprehenders' mean score (4.02, $SD = 0.56$) was significantly higher than the high-comprehenders' mean score (3.56, $SD = 0.79$) on the 'speech rate' sub-scale, $t(83) = 3.178$, $p = 0.002$. There was no difference between the high- and low-comprehenders' scores on the three other sub-scales. Thus, these results suggest that the test takers with lower comprehension scores in the unscripted group might have perceived the unscripted text as having slang and colloquial language because there was a good amount of vocabulary they could not decipher. Similarly, the low-comprehenders might have attributed their inability to comprehend the spoken texts to the seemingly rapid speech rate. These conclusions are merely speculative, however. Because there is so little research on the extent to which L2 listeners can perceive the organization, phonology, and fluency characteristics of unscripted spoken texts, more work is obviously needed in this area.

Implications and Conclusion

It is almost universally acknowledged that the goal for adult L2 learners in regard to their own pronunciation is intelligibility (e.g. Ballard & Winke, this volume; Harding, this volume; Isaacs, 2013; Isaacs & Trofimovich, 2012; Trofimovich & Isaacs, this volume), given that a fully 'nativelike' pronunciation is usually an unrealistic, inappropriate expectation for adult learners. This belief is confirmed by the fact that 'sounds like a native speaker' is no longer used as a descriptor on pronunciation rubrics/ratings scales. Similarly, it is almost universally acknowledged that real-world spoken language contains connected speech and hesitation phenomena that are not the result of 'lazy' or 'sloppy' pronunciation, but are in fact a normal, necessary and appropriate result of articulating spontaneous spoken language. And yet, believing that they are making listening comprehension more accessible to learners by maximizing intelligibility, L2 materials and test developers continue using unrealistic and inauthentic models of pronunciation in the spoken texts in their materials. Clark (2014) has demonstrated that it is feasible to commission semi-scripted spoken texts for L2 listening tasks, yet the vast majority of L2 listening tests use spoken texts with pronunciation involving formal, over-enunciated citation forms of language that differ dramatically from spontaneous, real-world spoken language. Indeed, by perpetuating inauthentic speech models at the expense of appropriate models of real-world speech, we believe that learners are disadvantaged in that they

acquire inaccurate perceptions of what L2 speakers should sound like and consequently feel unprepared to engage in discourse beyond the classroom. As in Wagner (2014) and Wagner and Toth (2014), we suggest that simplification strategies other than text modification be used to make real-world L2 speech accessible and intelligible to learners, including the careful management of: (a) text length; (b) the targets of attentional focus; (c) the intended depth of learners' comprehension; (d) the number of listening rounds; and (e) opportunities for hypothesizing, feedback and knowledge consolidation. Ultimately, teachers must help learners cope with their inability to understand *everything*, so that learners can build confidence in their ability to understand *something* and thus establish a grounding in comprehension that will sustain them in natural conversation. Similarly, drawing learners' attention to the characteristics of unplanned spoken language should also help learners notice and attend to these characteristics in subsequent spoken input, both inside and outside the language classroom, which corresponds very closely with Vandergrift and Goh's (2012) metacognitive approach to L2 listening instruction.

This study has demonstrated that L2 learners can identify spoken texts that are specially created for L2 learners, and even distinguish the organization, phonology and fluency characteristics of these texts from spontaneous, real-world spoken language. L2 test developers must therefore consider aspects of pronunciation not only when developing speaking tests, but also when developing L2 listening tests. They should include unscripted spoken texts in L2 listening tests, because doing so will result in better domain coverage (i.e. texts that are more reflective of the spoken texts in the real world) and more valid inferences about test takers' ability to understand real-world spoken language. In addition, the inclusion of these types of spoken texts can have a positive washback effect not only on test takers, but also for the larger educational systems that prepare students to take them, by promoting the use of unscripted, spontaneous texts in the L2 classroom and materials. If students know that these types of texts will appear in L2 listening tests, then they should be more receptive to their use in the classroom, even if they perceive them as initially more difficult. Similarly, L2 teachers and curriculum and materials developers should regularly implement unscripted spoken texts in L2 listening tasks, especially with more advanced learners, but even beginning learners can benefit from being exposed to these types of texts. As our results suggest, learners can readily tell when they are hearing inauthentic speech. If indeed the possibility of engaging with real-world spoken language strengthens learner motivation, as Peacock (1997) suggests, while also provoking anxiety, then our primary instructional concern should be providing sufficient support during experiences of real-world texts to make comprehension possible and thereby build among learners a repertoire of successful experiences that ultimately leads to a noticing of and familiarity with unscripted, spontaneous communication.

References

Bachman, L. and Palmer, A. (1996) *Language Testing in Practice*. Oxford: Oxford University Press.

Biber, D. (1988) *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D. (2006) *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

Biber, D. and Gray, B. (2013) Discourse characteristics of writing and speaking tasks types on the TOEFL iBT Test: A lexico-grammatical analysis. TOEFL Research Report No. RR-13-04. Princeton, NJ: Educational Testing Service.

Blau, E. (1990) The effect of syntax, speed, and pauses on listening comprehension. *TESOL Quarterly* 24, 746–753.

Brown, G. (1995) *Speakers, Listeners, and Communication*. Cambridge: Cambridge University Press.

Brown, J.D. (ed.) (2012) *New Ways in Teaching Connected Speech*. Alexandria, VA: TESOL International Association.

Brown, J.D. and Kondo-Brown, K. (2006) Introducing connected speech. In J.D. Brown and K. Kondo-Brown (eds) *Perspectives on Teaching Connected Speech to Second Language Speakers* (pp. 1–16). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.

Buck, G. (2001) *Assessing Listening*. Cambridge: Cambridge University Press.

Chafe, W. (1982) Integration and involvement in speaking, writing, and oral literature. In D. Tannen (ed.) *Spoken and Written Language: Exploring Orality and Literacy* (pp. 35–53). Norwood NJ: Ablex.

Chafe, W. (1985) Linguistic differences produced by differences between speaking and writing. In D. Olson, D. Torrance and A. Hildyard (eds) *Literacy Language and Learning* (pp. 105–123). Cambridge: Cambridge University Press.

Clark, M. (2014) The use of semi-scripted speech in a listening placement test for university students. *Papers in Language Testing and Assessment* 3 (2), 1–26.

Cucchiarini, C., van Doremalen, J. and Strik, H. (2010) Fluency in non-native read and spontaneous speech. *Proceedings of Disfluency in Spontaneous Speech (DiSS) and Linguistic Patterns in Spontaneous Speech (LPSS) Joint Workshop* (pp. 15–18).

Dupuy, B. (1999) Narrow listening: An alternative way to develop and enhance listening comprehension in students of French as a foreign language. *System* 27, 351–361.

Field, J. (2008) *Listening in the Language Classroom*. Cambridge: Cambridge University Press.

Flowerdew, J. and Miller, L. (1997) The teaching of academic listening comprehension and the question of authenticity. *English for Specific Purposes* 16, 27–46.

Freedle, R. and Kostin, I. (1999) Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing* 16, 2–32.

Gallien, C., Hotho, S. and Staines, H. (2000) The impact of input modifications on listening comprehension: A study of learner perceptions. *JALT Journal* 22, 271–295.

Gilmore, A. (2004) A comparison of textbook and authentic interactions. *ELT Journal* 58, 363–374.

Gilmore, A. (2007) Authentic materials and authenticity in foreign language learning. *Language Teaching* 40, 97–118.

Griffiths, R. (1991) The paradox of comprehensible input: Hesitation phenomena in L2 teacher talk. *JALT Journal* 13, 23–41.

Griffiths, R. (1992) Speech rate and listening comprehension: Further evidence of the relationship. *TESOL Quarterly* 26, 385–391.

Guariento, W. and Morley, J. (2001) Text and task authenticity in the EFL classroom. *ELT Journal* 55, 347–353.

Halliday, M.A.K. (1985) *Introduction to Functional Grammar*. London: Edward Arnold.

Haviland, S. and Clark, H. (1974) What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior* 13, 512–521.

Isaacs, T. (2013) Pronunciation. In *Cambridge English Centenary Symposium on Speaking Assessment* (pp. 13–15). Cambridge: Cambridge English Language Assessment.

Isaacs, T. (2014) Assessing pronunciation. In A. Kunnan (ed.) *Companion to Language Assessment* (Vol. 1, pp. 140–155). Oxford: Wiley-Blackwell.

Isaacs, T. and Trofimovich, P. (2012) Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition* 34, 475–505.

Ito, Y. (2006) The significance of reduced forms in L2 pedagogy. In J.D. Brown and K. Kondo-Brown (eds) *Perspectives on Teaching Connected Speech to Second Language Speakers* (pp. 17–26). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.

Kelch, K. (1985) Modified input as an aid to comprehension. *Studies in Second Language Acquisition* 7 (1), 81–90.

Kmiecik, K. and Barkhuizen, G. (2006) Learner attitudes towards authentic and specially prepared listening materials: A mixed message? *TESOLANZ Journal* 14, 1–15.

Ladefoged, P. and Johnson, K. (2010) *A Course in Phonetics* (6th edn). Boston, MD: Cengage.

McCarthy, M. (2005) Fluency and confluence: What fluent speakers do. *The Language Teacher* 29, 26–28.

McCarthy, M. (2010) Spoken fluency revisited. *English Profile Journal* 1, 1–15.

McCarthy, M. and Carter, R. (1995) Spoken grammar: What is it and how can we teach it? *ELT Journal* 49, 207–218.

McCarthy, M. and Carter, R. (2001) Ten criteria for a spoken grammar. In E. Hinkel and S. Fotos (eds) *New Perspectives on Grammar Teaching in Second Language Classrooms* (pp. 51–75). Mahwah, NJ: Lawrence Erlbaum.

Meinardi, M. (2009) Speed bumps for authentic listening material. *ReCALL* 21, 302–318.

Messick, S. (1989) Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18 (2), 5–11.

Messick, S. (1996) Validity and washback in language testing. *ETS Research Report Series* 1, 1–18.

Parker, K. and Chaudron, C. (1987) The effects of linguistic simplification and elaborative modifications on L2 comprehension. *University of Hawai'i Working Papers in ESL* 6, 107–133.

Peacock, M. (1997) The effect of authentic materials on the motivation of EFL learners. *ELT Journal* 51, 144–156.

Richards, J. (2006) Materials development and research: Making the connection. *RELC Journal* 37, 5–26.

Rost, M. (2011) *Teaching and Researching Listening* (2nd edn). Harlow: Pearson Education.

Thompson, S. (2003) Text-structuring metadiscourse, intonation, and the signaling of organization in academic lectures. *Journal of English for Academic Purposes* 2, 5–20.

Vandergrift, L. and Goh, C. (2012) *Teaching and Learning Second Language Listening: Metacognition in Action*. New York: Routledge.

Voss, B. (1979) Hesitation phenomena as sources of perceptual errors for non-native speakers. *Language and Speech* 22, 129–144.

Wagner, E. (2010) Survey research in applied linguistics. In B. Paltridge and A. Phakiti (eds) *Continuum Companion to Second Language Research Methods* (pp. 22–38). London: Continuum.

Wagner, E. (2013a) Surveys. In C. Chapelle (ed.) *The Encyclopedia of Applied Linguistics* (pp. 5470–5475). Oxford: Wiley-Blackwell.

Wagner, E. (2013b) Assessing listening. In A. Kunan (ed.) *Companion to Language Assessment*. Oxford: Wiley-Blackwell.

Wagner, E. (2014) Using unscripted spoken texts in the teaching of second language listening. *TESOL Journal* 5, 288–311.

Wagner, E. and Toth, P. (2014) Teaching and testing L2 Spanish listening using scripted versus unscripted texts. *Foreign Language Annals* 47, 404–422.

Wagner, E. and Wagner, S. (2016) Scripted and unscripted spoken texts used in listening tasks on high stakes tests in China, Japan, and Taiwan. In V. Aryadoust and J. Fox (eds) *Current Trends in Language Testing in the Pacific Rim and the Middle East: Policies, Analyses, and Diagnoses* (pp. 103–123). Newcastle upon Tyne: Cambridge Scholars Publishing.

Zhao, Y. (1997) The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics* 18, 49–68.

Appendix: Post-test Questionnaire

Authentic versus modified sub-scale

(1) The spoken texts were **not** authentic; they were specially created for students learning Spanish.
(2) The texts used authentic spoken input, like is found in real life.
(3) The speakers planned and practised what they were going to say and read from transcripts.
(4) The texts required me to listen to authentic spoken language, the same type of spoken language that is found in real life.
(5) The texts that were used did **not** have authentic spoken input.

Pronunciation sub-scale

(1) The speakers spoke clearly and used very clear pronunciation, which made it easier to understand them.
(2) The speakers pronounced each word clearly and distinctly.
(3) It was hard to understand the speakers because they did **not** enunciate well and did **not** speak clearly.
(4) The speakers' pronunciation in the texts was similar to native Spanish speakers' pronunciation in real-life conversations (excluded from analyses due to low item-total correlation of this item with the other items in the pronunciation sub-scale).

Speech rate sub-scale

(1) The speakers on the spoken texts spoke quickly, the same rate that native speakers normally use with each other.
(2) The speakers in the spoken texts spoke slowly and enunciated each word.
(3) The speakers spoke slowly and clearly so that the listeners would be able to understand them.
(4) The speakers spoke quickly, at the same rate as native speakers in real-life conversations.

Use of slang sub-scale

(1) The speakers used only formal, standard Spanish, with **no** slang.
(2) The spoken texts often had slang and colloquial speech in them.
(3) The speakers did **not** use slang or informal expressions when they were speaking.
(4) The speakers used slang and informal expressions that are found in real-life language.

Pauses and false starts sub-scale

(1) The speakers often had a lot of pauses, fillers (things like 'eh …', 'em …', 'este …', 'tú sabes …'), and false starts in their speech.
(2) Because the speakers planned what they were going to say and read from a transcript, there were few pauses, false starts, and fillers (things like 'eh …', 'em …', 'este …', 'tú sabes …') in the spoken texts.
(3) I could tell that the speakers were reading from a transcript, because they did **not** have any pauses, false starts or mistakes in their speech.
(4) There were pauses, fillers and false starts in the texts, suggesting that the speakers did **not** plan what they were going to say, and were **not** reading from transcripts.