**THE JOURNAL OF TEACHING ENGLISH FOR SPECIFIC AND ACADEMIC PURPOSES**
Vol. 1, N° 2, 2013, pp. 67−92

UDC (811.111'276.6+378.147):347.176.2

# HELPING INTERNATIONAL TEACHING ASSISTANTS ACQUIRE DISCOURSE INTONATION: EXPLICIT AND IMPLICITL2 KNOWLEDGE

Greta J. Gorsuch

Classical and Modern Languages and Literatures, Applied Linguistics, Texas Tech University, Box 42071, Lubbock, Texas, 79409-2071, U.S.A.
Phone: (806) 742-3145, E-Mail: GRETA.GORSUCH@ttu.edu

**Abstract**. *This study explored a theoretically-driven permutation of an intervention designed to improve ITAs' spoken Discourse Intonation (DI). The object was to learn if implicit knowledge growth in DI could be found as the result of an experimental group participating in explicit instruction and in audio-assisted repeated reading treatments using twice-weekly easy, popular science texts for 14 weeks. In a read-aloud condition where speech processing burdens were reduced, both an experimental and control group (who received explicit instruction only) improved over time on speech rate, planning pauses versus hesitation pauses, prominence, tone choices, and length of tone choice pause groups. In a free-response task where processing burdens were increased, however, there was little evidence of change in implicit knowledge of DI for the experimental group. One positive thing was learned: Explicit DI instruction did not reduce participants' speech rate and thus participants could focus on form as well as meaning in extended speech. Explicit DI instruction, where form is linked to meaning, is worthwhile in that explicit knowledge may become proceduralized and available for learners' extemporaneous use. Implicit knowledge building in DI, while difficult to demonstrate, may still be worthwhile if it builds learners' knowledge of vocabulary (to improve prominence) and builds their experience hearing DI features linked to meaning within extended texts.*

**Key words**: *discourse intonation, L2 prosody, international teaching assistants, L2 acquisition, pedagogical intervention*

## 1. INTRODUCTION

International Teaching Assistants (ITAs) are (largely) Chinese, Indian, and Korean chemistry, math, etc., graduate students in the U.S. and Canada who are supported by teaching undergraduate courses in their areas of study (Ford, Gappa, Wendorff and Wright 1991; Griffee, Gorsuch, Britton and Clardy 2009; Smith, Byrd, Nelson, Barrett and Constantinides 1992). In 2010, their numbers increased to 242,061 in the U.S. and they comprised 15.5% of all graduate students (Council of Graduate Schools 2010). Many undergraduate science and math courses are taught by ITAs (Chiang, 2009; Gorsuch and Sokolowski 2007; Kaufman and Brownworth 2006). Indeed, ITAs make undergraduate education possible (Gorsuch 2011a; Williams 1992) and as instructors are contributors to the learning of undergraduates (McKeachie 2004).

ITAs must teach using English, their L2 (Cardillo 2002; Gorsuch 2011a; Kaufman and Brownworth 2006). While many ITAs have developed some control over L2 lexis and syntax and may be considered high intermediate to advanced learners, they have little experience using their spoken language in extended talk for social and instructional purposes (Gorsuch 2011a). Thus ITAs struggle with spoken fluency, and in particular, prosody.

One more recent focus of ITA need has been Discourse Intonation (DI), which is the use of pause groups (associated with fluency), and sentence-level stress and tone choices (associated with prosody) for communicative purposes (Brazil 1997). DI is used to emphasize and differentiate ideas, begin and end topics, and express social relationships (Pickering, 2001). Accumulating evidence suggests that the ability to use DI (prosody) is essential for ITAs' success as college-level instructors (Anderson-Hsieh and Koehler 1988; Gorsuch, Meyers, Pickering and Griffee 2010 2013; Hahn 2004; Pickering 1999; Wennerstrom 2000), yet for a variety of reasons DI is acquired late, and with difficulty (Gorsuch, 2011a 2011b; Hirst andDiCristo 1998; Pennington and Ellis, 2000). As use of appropriate DI may be considered a basis for listener perceptions of spoken communication ability (Butcher 1980; Ejzenberg 2000; Olynak, Anglejan and Sankoff 1990; Wennerstrom 2000), it is essential to devise and test pedagogical interventions which may bring about improvements in learners' DI (Gorsuch 2011b; see also Pickering 1999; Wennerstrom 2001).

This study explores a theoretically-driven permutation of an intervention designed to improve ITAs' DI. In 27 treatments spanning 14 weeks, seven ITAs in an experimental group engaged in combined awareness-raising tasks and audio-assisted repeated reading treatments designed to build up their implicit knowledge of DI through input. They also had 14 weeks of explicit DI instruction. For comparison, seven ITAs who were newly arrived in the U.S. formed a control group. While they had three weeks of intensive, explicit DI instruction, they did not have the input treatments, nor did they have extensive opportunities to hear continuous English speech in their environment. Changes in both groups' use of DI in audio recorded and transcribed parallel read-aloud and free response pre- and post-test tasks were observed.The read-aloud task was designed to track changes in participants' explicit DI knowledge, while the free-response task was designed to detect changes, if any, in participants' implicit DI knowledge.

## 2. Literature Review

This literature review begins with a description of the challenges for learning Discourse Intonation (DI), and continues with an argument for using instructional programs designed to develop both implicit and explicit knowledge for learners. The review concludes with detailed descriptions of aspects of DI that were highlighted in this study, how these features were operationalized, and predictions about participants' DI use in read-aloud versus free-response tasks, and over time.

### 2.1. Learning issues for discourse intonation

DI is defined as how a speaker uses the pausing and prosodic system of a language for communicative purposes, including pauses (pause groups), sentence level stress (prominence), and intonation contours (tone choices) (Brazil 1997). There is general agreement that L1 transfer is an issue in accounting for and learning DI and L2 phonology (Pennington 1996; Pickering 1999; Strange and Shafer 2008; Wennerstrom

2001; Zampini 2008), including the pausing patterns (pause groups) that are key to listeners' perceptions of fluency (Chambers 1997; Freed 1995). Learners' L1s affect whether and how DI features in English are perceived in aural input (Doughty 2003; Escudero 2007; Pennington and Ellis 2000). Pennington and Ellis (2000) and Wennerstrom (2001) speculate on additional effects due to English L2 learners' previous educational experiences, where text-based, accuracy-focused instruction results in choppy sounding word-for-word speech.

Greater acquisition of appropriate DI use has been associated with greater experience with L2 (Escudero 2007; Pickering 1999; Zampini 2008). This makes sense when one takes into account the attentional resources needed for L2 speaking. Many L2 learners do not have sufficient automaticity in grammatical encoding to plan utterances as semantic units (speaking in pause groups) (Chambers 1997; Dechert 1984; Deese 1980; Ejzenberg 2000; Foster, Tonkyn, and Wigglesworth 2000; Kormos and Denes 2004; Pickering 1999; Tavakoli 2010), or to use appropriate prominence or tone choices (Pickering 1999). When reading prepared texts aloud or retelling the same story repeatedly to different listeners, learners' processing load decreases, and they can then focus on using appropriate pause groups (Arevart and Nation 1991; Gibson 2008; Tench 1996). In other words, there is likely an effect due to task type (reading a text aloud versus unrehearsed speech or conversation; see Tables 1, 2, and 3 below).

Some broad developmental stages have been suggested where L2 learners either "have" DI or "do not have" it (e.g., Pickering 1999; Wennerstrom 1998 2001). For instance, "do not have DI" L2 learners were found to "have heavy stress on every word and consistent sentence-final pitch accent assignment" (Wennerstrom 2001, 247). Freed (1995), in studying French L2 learners abroad, found that as they gained experience using French, they used more fluent-sounding ways of hesitating (such as drawls), and also increased their speech rate. Tavakoli (2010, 77) posited that as learners' grammatical processing becomes proceduralized, "L2 speakers speak in longer units and therefore pause less."

While detailed evidence for developmental stages for acquisition of DI is scant, there is much evidence for idiosyncracy in acquiring these L2 features (Freed 1995; Wennerstrom 2001). Freed had difficulty making group comparisons (study abroad versus no study abroad) due to "extreme variability of scores within each group" (1995, p. 137). Both Wennerstrom (2001) and Levis, Muller Levis and Slater (2012) noted idiosyncratic variations in use of L2 (English) prominence within groups of Chinese ITAs. This may have been due to within-group variations of their overall English ability. Indeed, Wennerstrom (1998, 20) speculated that acquisition of prosody "takes place along with the development of other language skills, including grammar, vocabulary." Indian ITAs in Levis et al (2012) showed a command of lexico-grammatical structures equal to that of American ITAs, and while their use of prominence was not appropriate by North American standards, they showed less within-group variation.

Scholars in the field describe prosody (DI) as being an abstract system, and hard for learners to apply to specific instances of language use (Pennington and Ellis 2000; Zampini 2008). And in any event, prosody is not often taught consistently, nor in ways that are connected to meaning (Chambers 1997; Gilbert 1995; Pickering 1999). To complicate matters, prosody is hard to hear in input (e.g., Doughty 2003; Kiany and Shiramimy 2002), with L2 learners using other cues for getting meaning from what they hear (Pennington and Ellis 2000).

### 2.2. An argument for developing explicit and implicit knowledge for DI learners

Given the importance of appropriate prosody for effective spoken communication (Munro 2008), DI should be explicitly taught. Over time, with practice and feedback, explicit knowledge may become proceduralized and available for use in talk. At the same time, the "default processing mode" for second language acquisition processes "is implicit" (Doughty 2003, 292, emphasis in the original). To use DI, which is arguably a complex, resource-demanding process, requires complex knowledge. Such knowledge, according to Doughty is acquired as "implicit knowledge leading directly to procedural ability" (2003, 291). This suggests that intensive treatments of comprehensible input may be useful, particularly if designed to highlight pause groups, prominence, and tone choices in continuous and extended talk. In the current study, the twice-a-week awareness-raising tasks and audio-supported repeated reading treatments were designed to reinforce participants' declarative knowledge learned through direct instruction in class sessions, but also to develop participants' implicit knowledge (e.g. Abe 2009; B. VanPatten personal communication November 7, 2009).

Pre- and post-test measures of a read-aloud task and a free-response task may give clues as to whether experimental and control group participants can apply explicit and/or implicit knowledge to two tasks that would demand high or low levels of attentional resources. The two different tasks may also reveal whether there are changes in learners' application of knowledge, and perhaps use of DI, over time.

### 2.3. Pause groups

Pause groups (also called thought groups) are utterances that represent a thought that has been encoded into a grammatical chunk (Butterworth 1980; Pawley and Syder 2000). Pause groups are bounded by pauses (Bolinger 1998; Crystal and Davy 1969; Goldman Eisler 1968; Levelt 1989) and are marked by a continuous prosodic contour (Pennington 1996). In native English speakers' speech, pauses of varying lengths usually appear at clause and sentence boundaries, giving listeners important clues as to whether an utterance or topic is complete. An example is can I ask a question // for the lab report // do I use the equation for each piece of data I observe // or do I just solve the equation once // to show I know it where // represents pauses of .3 seconds or longer.

Compared to native English speaking teaching assistants (TAs), ITAs pause more often and more erratically, tending "to regularly break up conceptual units" (Pickering 1999, 51-52; see also Anderson-Hsieh and Koehler 1988; Gorsuch 2011b), placing a greater burden on listeners (U.S. undergraduates). An example from a participant in this study is: um my project is // focused on // synthesis of // far right dye named the two SPBO. In addition to the frequent, long pauses which violate clause boundaries, the shortness of the pause groups suggests difficulty on the part of the ITA in planning and encoding what she wishes to say about her research, a topic well known to her (Olynak et al 1990; Schmidt 1992; see also Deese 1980; Lennon 2000; Riggenbach 1989).

In this study participants' pause groups were measured using the speech rate, fluent run, and unfilled pause measures shown in Table 1 below. Descriptions are given of how these features may vary between read-aloud tasks and free-response tasks, and also how learners with changing levels of explicit or implicit knowledge may use the features differently over time (explicit instruction to improve declarative knowledge versus explicit + input + awareness raising instruction to develop implicit knowledge).

Table 1 Measures for pause groups

| Measure | Possible task effects, developmental changes | How done | Example | Sources |
|---|---|---|---|---|
| Speech rate | **Task effects**: In read-aloud tasks, participants' speech rate will likely be faster. In free response tasks, the speech rate will be slower due to increased processing demands. **Developmental changes**: When we accentuate second language forms (such as DI) through explicit instruction, learners' limited attentional resources will be consumed by this focus, resulting in slower rates of speech (less fluency) (Ellis, 1997). If participants use explicit knowledge their speech rates may remain low across task types. If their implicit knowledge is developed, their speech rates may be higher in free-response tasks. | Counting the number of meaningful syllables uttered per minute including repetitions, self-corrections, and recognizable parts of words. | Pause group *um it's a it's a blue color dye we should use it on the* = 14 syllables | Chambers, 1997; Skehan, 2003 |
| Length of "fluent runs" | **Task effects**: Fluent "runs" of participants reading a text aloud will be longer; participants are freed of having to plan their own speech. In a free response task, fluent "runs" will be shorter as processing demands increase. **Developmental changes:** If participants learn through explicit instruction their fluent "runs" may remain low (e.g., Freed, 1995), particularly in free response tasks. If their implicit knowledge is developed, their fluent runs may increase over time, particularly in free response tasks. | Counting the number of meaningful syllables between pauses of .3 seconds or more and then dividing by the total number of pause groups used by the participant | *OK ah* // *now I'm doing the research with ah* // *aroptopsis* // *uh* // *it's a* // *kind of a model plant* = 22 syllables divided by six pause groups = 3.67 syllables per pause group | Chambers, 1997; Kormos & Demes, 2004; Skehan, 2003 |

| | | | | |
|---|---|---|---|---|
| Unfilled planning and hesitation pauses per minute | **Task effects:** Participants will use more unfilled planning pauses than unfilled hesitation pauses per minute in a read aloud task due to reduced processing demands. In free response tasks, participants will use more unfilled hesitation pauses than planning pauses as they struggle with encoding their own speech. **Developmental changes**: If participants learn through explicit DI instruction they may improve over time in read-aloud tasks, as they may choose as their main challenge to render the script into grammatical chunks (more planning pauses). This may mean their explicit knowledge is proceduralized. If their implicit knowledge is developed, participants may use more unfilled planning pauses over time, particularly in free-response tasks. | **All unfilled pauses:** Silences of greater than .3 seconds were marked (//) on transcripts. **Unfilled planning pauses:** Pauses coinciding with complete clauses or sentences were counted and calculated as a per-minute measure (number of pauses divided by seconds and multiplied by 60) **Unfilled hesitation pauses**: Pauses occurring in the middle of complete clauses or sentences were counted and calculated as a per-minute measure. (number of pauses divided by seconds and multiplied by 60) | *my advisor don't want to push me very hard // uh I just do some simple // experiments in order to familiarize with the instruments and // right now uh //* The first pause group ends in one planning pause, whereas the second, third, and fourth pause groups are interspersed by three hesitation pauses. | Chambers, 1997; Kormos & Demes, 2004 |

### 2.4. Prominence

Prominence is the stressing of key words in utterances using higher pitch, longer vowels, and greater volume (Pennington 1996; Pickering 2010). Speakers use prominence to highlight information they believe is important within and between thought groups (Wennerstrom 1998 2001). This might be information that the speaker believes is new to listeners, or ideas that need to be contrasted as in miTOsis cells are simply REplicated // meiOsis on the other hand // is needed for SExual reproduction (e.g., Pickering 1999;Wennerstrom 1998). Prominence is a critical feature of academic talks in which listeners must parse information, and comprehend and predict the content of messages (Anderson-Hsieh and Koehler 1988; Hahn 2004; Pickering 1999).

Accumulating evidence suggests that many ITAs do not use prominence, either speaking in a monotone or emphasized words unsystematically without regard to their information value (Levis et al 2012; Pickering 1999; Wennerstrom 2001). Undergraduates listening to talks given by ITAs got higher lecture comprehension scores where ITAs used more native-like prominence patterns (Wennerstrom 2001).

In this study participants' use of prominence was measured using the prominent syllable and unjustified prominence measures shown in Table 2 below.

Table 2 Measures for prominence

| Measure | Possible task effects, developmental changes | How done | Example | Source |
|---|---|---|---|---|
| Prominent syllables per minute | **Task effects**: Participants may use more prominent syllables in a read-aloud task due to reduced processing demands. In a free-response task, participants may sound more monotone (using fewer prominent syllables) due to struggles with speech planning. **Developmental changes:** Participants with explicit instruction may have some proceduralized explicit knowledge and use more prominent syllables, particularly in a read-aloud task. Participants with more implicit knowledge may use more prominent syllables in a free-response task. | Counting the number of syllables or one-syllable words that were substantially louder, or higher in pitch than other syllables in the speech sample, as judged visually from a decibel or pitch waveform. Counts were then calculated as a per-minute rate (number of all prominent syllables divided by seconds and multiplied by 60) | *eh eLEtrochemisty // um generally speaking is a STUdy of battery* = two prominent syllables | Developed for this study. |
| Unjustified prominent syllables per minute | **Task effects:** Participants may use fewer unjustified prominent syllables in a read-aloud task due to reduced processing demands (more systematic in their use of prominence). In a free-response task, participants may use more unjustified prominent syllables (be less systematic) due to struggles with speech planning. **Developmental changes**: Participants with explicit instruction may have more or less proceduralized knowledge and thus use fewer OR MORE unjustified prominent syllables (due to overgeneraliza-tion) in either read-aloud and free-response tasks. Participants with more implicit knowledge may use fewer unjustified prominent syllables in a free-response task. | Counting the number of prominent syllables in fully formed words which seemed unmotivated (no new or important information was being given, nor transitions, nor comparisons). Counts were then calculated as a per-minute rate (number of unjustified prominent syllables divided by seconds and multiplied by 60) | *the evaporation OF water is a PHYsical change OF* seems unmotivated by its context. | Concept adapted from Wennerstrom, 2001. |

### 2.5. Tone choices

Tone choices are expressed as noticeable and measurable shifts in pitch which are either rising, level, or falling (Gorsuch et al 2010 2013), generally at thought group boundaries (Wennerstrom 1998). Tone choices are used to indicate whether a speaker has finished speaking (falling tones) or plans to continue speaking (rising and level tones) (Gorsuch et al 2010 2013), and whether a speaker wishes to begin, continue, or end a topic (Wennerstrom 1998 2001). Finally, tone choices are key to establishing rapport between interactants (Pickering 1999 2010). A "variety in pitch...signals interest or excitement in the topic" (Pennington 1996, 160), whereas a speaker who uses only level or falling tones sounds unfriendly, bored, or angry, as in about the exPERiment // your SETup //really simple versus about the exPERiment // your SETup //really simple.

Unfortunately, ITAs overuse level and falling tones (Pickering 1999), even when intending to continue a topic (Wennerstrom 1998; see also Lennon 2000; Young 1994). Pickering suggested that ITAs “are unaware of some of the social and informational functions of tone choices” used by native English speakers around them (1999, 169). Overuse of level or falling tones in non-native English speakers’ extended speech has also been attributed to linguistic processing problems expressed as hesitations (Pickering 1999, 169): “online verbal planning clearly accounts for many of the truncated tone units” (see also commentary by Field 2000 2008; Gorsuch 2011b; Wennerstrom 1998).

In this study participants’ use of tone choices was measured using the rising, level, and falling tone-per-minute, and syllables-per rising, level, and falling tone measures shown in Table 3 below.

Table 3 Measures for tone choices

| Measure | Possible task effects, developmental changes | How done | Example | Sources |
|---|---|---|---|---|
| Rising, level, and falling tones per minute | **Task effect**: If participants have explicit knowledge they may use more rising and falling tone choices (as opposed to predominantly level tones with some falling tones) in a read-aloud task. In a free-response task with higher planning and processing demands, participants may use fewer rising and falling tones and revert back to predominantly level with some falling tones. **Developmental changes**: Participants with explicit knowledge may use a greater variety of rising, level, and falling tones more over time, particularly in read-aloud tasks, suggesting more proceduralized knowledge. Participants who have more developed implicit knowledge may use a variety of tone choices in both read-aloud and free-response tasks. | Pause groups were assigned a category as having either a rising, level, or falling tone. Each type of pause group was counted. These counts were then calculated as per-minute rates (number of tone choice type divided be seconds and multiplied by 60). | *so but the battery is very➔ // different from➔ // the general batteries➘// used in ➔// um ➔// TV control or recorder or cell phone➔* This sample has five pause groups with level tones and one pause group with a falling tone. | Developed for this study. |

| | | | | |
|---|---|---|---|---|
| Number of syllables per rising, level, and falling tone pause group | **Task effect**: If participants have explicit knowledge of tone choices, they may be able to encode longer rising and falling tone choice pause groups (more syllables), along with level tone pause groups in a read-aloud task. With more processing demands in a free-response task, participants may be unable to pay attention to tone choices at the discourse level, and will encode shorter rising and falling tone choice pause groups overall (fewer syllables).<br>**Developmental changes**: Participants who use explicit knowledge may be able to use longer utterances of any tone choice type in pragmatically meaningful ways in a read-aloud task. But in a free-response task they might have fewer syllables on average for any type of tone choice pause group, as in *and a ➔ // with a lot of uses ➘* Participants with more developed implicit knowledge may be more successful in encoding longer runs of speech embedded in pragmatically meaningful tone choice pause groups. | Meaningful syllables per rising, level, and falling tone choice pause groups were counted. The counts were calculated as averages for each pause group type. | *currently currently I am ➔ // doing some research about ➔ // organic chemistry ➘ // but um actually I am the student ➔ // whose division is in analytical chemistry ➘// so my ➚ // program is um a little different from other graduate students ➘* There are 2 meaningful syllables in one rising tone pause group, thus the average for this sample is 2 syllables per rising tone pause group.<br>There are 25 meaningful syllables in three level tone pause groups, for an average of 8.33. There are 38 meaningful syllables in three falling tone pause groups for an average of 12.67 syllables. | Developed for this study but inspired by Chambers, 1997; Kormos & Demes, 2004; Skehan, 2003 |

3. RESEARCH QUESTIONS

This descriptive study compares which, if any, aspects of an experimental group's and control group's Discourse Intonation (DI) seem to change over time in a read-aloud task and a free response task. Specifically:

1. What are participants' speech rates, length of fluent runs, and unfilled planning and hesitation pause rates on read-aloud and free-response tasks at the beginning and end of two types of instruction: instruction aimed at building implicit and explicit DI knowledge (experimental group) versus instruction aimed only at building explicit DI knowledge (control group)?

2. What are participants' prominent syllable-per-minute rates and unjustified prominent syllable-per-minute rates on read-aloud and free-response tasks at the beginning and end of the two types of instruction?

3. What are participants' rising, level, and falling tone-per-minute rates and number of syllables per rising, level, and falling tone pause groups on read-aloud and free-response tasks at the beginning and end of the two types of instruction?

## 4. METHOD

### 4.1. Participants

The experimental group (n = 7) was comprised of five males and two females, in their early to mid-twenties. All had Mandarin as their L1. They were seeking degrees in math, biology, and chemistry. They had been at the institution where the study took place an average of 8.43 months and were enrolled in a semester-long classroom communication ESL course. The group was enrolled in the course because they had failed to be approved to teach in a previous summer workshop or a semester-long course on the basis of listening and spoken performance tests. By the time they began the course and treatment their SPEAK test scores (n = 4) only averaged 40, which is considered to be a level where a speaker can communicate only somewhat effectively (Educational Testing Service, 1996). Three of the students had TOEFL ibt speaking scores with an average of 17 considered to be "fair." While the participants had some control over grammar and a working implicit vocabulary, they were not fluent in their speech, even after five or so months of living and studying in the U.S. In the parlance of the program in which the course was offered, the participants had "stalled out" and were not improving. While there were other class members (n = 3) they were excluded from the data because they had Korean, Cantonese, and Persian as their L1s. L1 is considered to be a significant factor in learning Discourse Intonation (DI) and so this was an attempt to control this variable.

The control group (n = 7) had three males and four females, and they were seeking similar degrees (math, biology, chemistry, and computer science). They were attending a summer ITA workshop and had been at the institution for only a week when the pre-test data were collected. They had not previously spent time in the U.S. Their SPEAK test scores (n = 7) averaged 40 halfway through the three-week workshop. The control group was chosen at random from a pool of 20+ newly arrived workshop participants whose L1s were Mandarin. This meant they likely had few opportunities to hear continuously spoken English and thus had little access to the input which might change their implicit knowledge of DI.

### 4.2. Materials

#### *4.2.1. Experimental and control group explicit DI instruction*

Both experimental and control groups used the same materials to build explicit knowledge of DI, including thought groups (pausing patterns in English), prominence (sentence-level stress), and tone choices (intonation) by linking specific forms to meaning (e.g., using falling tones is associated with ending an idea). The explicit instruction consisted of short listening tasks, and inductive and deductive sessions based on authentic classroom recordings and their transcripts. Participants in both groups engaged in rehearsed and free speaking practice, and feedback sessions in pairs and groups, including teaching simulations. There were differences in the intensity and duration of the explicit instruction between the two groups: The experimental group had, twice-weekly, 90-minute classes for 14 weeks and the control group had five-hour classes, daily, for 2 ½ weeks.

#### *4.2.2. Experimental group implicit knowledge treatments*

The additional implicit knowledge treatments for the experimental group were comprised of 27 awareness-raising and audio-supported repeated reading (RR) sessions, each lasting 20-25 minutes. Each session was done by experimental group participants immediately before regular class meetings. Each session began with an awareness-raising listening task based on the RR text to be used in the subsequent part of the session. Using the awareness-raising tasks was motivated by research (Gorsuch 2011b) that ITAs' attention needed to be more reliably directed to DI features in aural input, in that participants might be too focused on comprehending the meaning of the extended, authentic discourse presented in the RR treatment texts, rather than noticing specific DI forms. Using commentary from Chun, Hardison, and Pennington (2008) and Pennington (1996), the awareness-raising tasks focused on short excerpts of the day's RR text and asked participants to focus on a specific DI form, such as this handout below on rising tones:

**Faraway planet could support life *By Stephen Ornes***

Draw ➚ in front of every pause group that ends in a rising tone.
_____ During a September 29 briefing for the press,
_____ the researchers described a planet that is probably just right.
_____ Called Gliese 581g,
_____ it is about three times as massive as Earth.
_____ It orbits its star so closely
_____ that it goes all the way around in only 37 days.
_____ Because it's so close,
_____ it's tidally locked,
_____ which means one side always faces its star,
_____ and the other side is always dark.
_____ Temperatures on Gliese 581g's surface
_____ probably range between -24 degrees
_____ and 10 degrees Fahrenheit.

Each awareness-raising task included a short debriefing session in which participants were asked to compare answers (they circled or wrote their answers on the handout) and then speculate on how the feature was related to the speaker's intended meaning (the intended meaning here was "I'm not finished").

Then the participants did an audio-supported RR treatment. The procedure was:

1. Participants silently read a segment of a 500-word popular science text once while timing themselves with a stop-watch, and wrote their time on a log sheet. Instructors briefly answered any questions on word meaning, pronunciation, or grammar point.
2. Participants then read the same text second and a third time while listening to an audio file of the text or hearing it read aloud by the teacher. Participants had the option to quietly speak along with the audio model.
3. Participants finally silently read the text a fourth time while timing themselves with a stop watch, and wrote their time on a log sheet.
4. Participants wrote a short report in their L1 or L2 - their choice.

The treatment texts were 350-600 word segments of popular science texts taken from the website Science for Kids (http://www.sciencenewsforkids.org/). The website texts were original pieces on science and were written at the junior and senior high school level. As estimated, using basic readability formulae, the texts for the first eight treatments averaged a 7.03 grade level, 8.3 grade level for the second eight texts, and so on. The point was to use texts that were easy enough to be comprehended, so the input from them would be comprehensible, but even more importantly, so the learners could focus on specific DI features in the audio model as they progressively built up experience with the text through the repetitions specified in the treatment procedure. In the post-treatment reports, experimental group participants indicated that texts were relatively easy, yet they also mentioned new words and sentence- and discourse-level textual features they had noticed. Their silent reading rates increased slightly as the semester went on, suggesting good comprehension (see Gorsuch and Taguchi 2008 2010).

*Read-aloud and free-response tasks.* The read-aloud text for both the pre- and post-test were paragraphs from freshman texts from the discipline each participant was seeking a degree in. Participants read the passage silently and then read it out loud onto an audio recorder. Here is an example chemistry passage: Every substance has a unique set of properties, or characteristics, that allow us to recognize it and to distinguish it from other substances. Properties of matter can be grouped into two categories: physical and chemical. Physical properties are those properties that we can measure without changing the basic identity of the substance. Participants read the same text for both the pre- and the post-test.

The free-response task was participants' extemporaneous response to the query "Tell me about your research." If participants spoke for less than one minute, the researcher asked a brief follow up question to get the participant talking again for at least 30 seconds. Participants' responses were audio recorded and they were asked the same question for both the pre- and the post-test.

#### *4.2.3. Read-aloud and free-response tasks*

The read-aloud text for both the pre- and the post-test were paragraphs from freshman texts from the discipline each participant was seeking a degree in. Participants read the passage silently and then read it out loud onto an audio recorder. Here is an example chemistry passage: Every substance has a unique set of properties, or characteristics, that allow us to recognize it and to distinguish it from other substances. Properties of matter can be grouped into two categories: physical and chemical. Physical properties are those properties that we can measure without changing the basic identity of the substance. Participants read the same text for both the pre- and post-test.

The free-response task was participants' extemporaneous response to the query "Tell me about your research." If participants spoke for less than one minute, the researcher asked a brief follow up question to get the participant talking again for at least 30 seconds. Participants' responses were audio recorded and they were asked the same question for both the pre- and post-test.

### 4.3. Procedure

At the beginning of the long-semester ITA course, experimental group participants were interviewed individually in the researcher's office. They were aware their talk was being recorded, and were taught how to turn off the recorder, if they wished to say something off the record. After warm-up questions, participants read the prepared text aloud and then responded to the free-response prompt. The same procedure was followed for the post-test at the end of the semester. For the control group the task procedure was the same, except the pre-test was done at the beginning of a three-week workshop and the post-test was done at the end. As a way to compare all participants' speech to that of the native speakers doing the same tasks, one native speaker of English (not the researcher) read aloud all of the passages used in the read-aloud task. A second native speaker responded to the free-response query, “Tell me about your research.”

All participants’ continuous talk was transcribed by someone not involved in the study. The audio files and each accompanying transcript were randomly assigned a code number so that the researcher did not know whether she was analysing the talk of an experimental or control group member, or a pre- or post-test. After all audio files and transcripts were analyzed for pause groups (Table 1), prominence (Table 2), and tone choices (Table 3), they were set aside for several months. Five files were randomly chosen for a re-check of the analyses as a means of conducting intrarater reliability. There were no differences between these analyses, suggesting consistency. The audio files and transcripts of the native English speakers were analyzed last, using the same measures.

### 4.4. Analyses

To answer RQ #1 on pause groups, RQ #2 on prominence, and RQ #3 on tone choices, the audio files and transcripts were analyzed using the measures described in Tables 1, 2 and 3. Descriptive statistics, including means and standard deviations, were calculated for each measure for each group (experimental versus control) on two different tasks (read-aloud versus free-response), and for each occasion (pre-test versus post-test). Because of the small number of participants and the large number of measures and comparisons statistical analyses were done.

## 5. Results

### 5.1. RQ #1 on pause groups

Descriptive statistics for speech rates, length of fluent runs, and unfilled planning and hesitation pauses per-minute on the read-aloud and free-response tasks are given in Tables 4 and 5.

Table 4 Descriptive statistics for speech rate, length of fluent runs, and unfilled planning and hesitation pauses on read-aloud task

| | Experimental Group | | Control Group | |
|---|---|---|---|---|
| | Pre-test | Post-test | Pre-test | Post-test |
| Speech rate (syllables/min.) | *M* = 111.59<br>*SD* = 46.71 | *M* = 156.78<br>*SD* = 73.62 | *M* = 206.96<br>*SD* = 22.59 | *M* = 216.78<br>*SD* = 14.59 |
| Length of fluent runs (syllables/pause group) | *M* = 8.27<br>*SD* = 1.7 | *M* = 9.64<br>*SD* = 2.42 | *M* = 10.47<br>*SD* = 3.69 | *M* = 12.03<br>*SD* = 2.59 |
| Unfilled planning pauses (per minute) | *M* = 10.68<br>*SD* = 5.57 | *M* = 14.09<br>*SD* = 6.84 | *M* = 11.94<br>*SD* = 5.75 | *M* = 15.63<br>*SD* = 4.04 |
| Unfilled hesitation pauses (per minute) | *M* = 3.55<br>*SD* = 3.21 | *M* = 2.48<br>*SD* = 1.93 | *M* = 9.27<br>*SD* = 8.2 | *M* = 3.00<br>*SD* = 3.03 |

Table 5 Descriptive statistics for speech rate, length of fluent runs, and unfilled planning and hesitation pauses on free-response task

| | Experimental Group | | Control Group | |
|---|---|---|---|---|
| | Pre-test | Post-test | Pre-test | Post-test |
| Speech rate (syllables per minute) | *M* = 85.49<br>*SD* = 23.85 | *M* = 97.5<br>*SD* = 15.19 | *M* = 146.34<br>*SD* = 39.45 | *M* = 152.75<br>*SD* = 39.49 |
| Length of fluent runs (syllables/pause group) | *M* = 5.55<br>*SD* = 1.33 | *M* = 5.45<br>*SD* = 1.24 | *M* = 5.56<br>*SD* = 1.64 | *M* = 5.96<br>*SD* = 2.12 |
| Unfilled planning pauses (per minute) | *M* = 4.55<br>*SD* = 2.45 | *M* = 3.90<br>*SD* = 1.39 | *M* = 5.71<br>*SD* = 2.50 | *M* = 5.49<br>*SD* = 1.67 |
| Unfilled hesitation pauses (per minute) | *M* = 11.08<br>*SD* = 2.92 | *M* = 14.2<br>*SD* = 3.84 | *M* = 20.53<br>*SD* = 6.19 | *M* = 21.5<br>*SD* = 6.46 |

### *5.1.2. Task effects on pause groups*

The results showed what appeared to be task effects, as predicted. Both groups had higher speech rates, longer fluent runs, more unfilled planning pauses and fewer unfilled hesitation pauses for the read-aloud task ("better" fluency performance). For the free-response task where the speech processing burden was higher, the fluent performance of both groups seemed worse. Most dramatic were the groups' pausing patterns. When both groups did the free-response task, their unfilled hesitation pauses outnumbered their planning pauses, suggesting less efficient encoding of grammatical chunks, and more choppy sounding and possibly hard-to-follow speech. On the pre-test read-aloud task the experimental group used M = 10.68 planning pauses per minute versus M = 3.55 hesitation pauses. Thus, only 25% of all pauses were hesitation pauses. On the free-response task, the experimental group used only M = 4.55 planning pauses per minute versus M = 11.08 hesitation pauses, meaning that 71% of all pauses were hesitation pauses. To put this difference in task performance into context, and to further demonstrate the strong effects of increasing processing burdens on non-native English speakers, the unfilled pausing data from native English speakers are given here: one native English speaker averaged 16.91 planning pauses per minute across the four readings on the read-aloud, and only 3.86 hesitation pauses, meaning that 19% of all pauses were hesitation pauses. On the free-response task, the second native speaker used 14.73 planning pauses per minute, and 6.69 hesitation pauses, meaning that only 31% of all pauses were hesitation pauses.

### *5.1.3. Developmental changes on pause groups*

As can be seen in Tables 4 and 5, the control group spoke faster on both read, read-aloud and free-response tasks on both the pre- and post-tests, suggesting a pre-existing difference between the groups at the outset of the study. This made interpretations of comparisons of developmental changes between groups harder to do. On the read-aloud pre-test, the experimental group spoke M = 111.59 syllables per minute and on the post-test increased to M = 156.78, while the control group spoke on average M = 206.96 syllables per minute on the pre-test, and increased, although a bit less, to M = 216.78. On the free-response task the experimental group spoke M = 84.49 syllables per minute on the pre-test, and increased to M = 97.15 on the post-test. The control group also increased their average speech rate from the pre-test M = 146.34 (syllables per minute) to M = 152.75 on the post-test.

Setting speech rate aside, both groups seemed less different at the outset of the study, particularly when it came to the free-response task. In terms of length of fluent runs, both groups seemed able to encode slightly longer pause groups from the pre-test to the post-test on the read-aloud task. On the pre-test, the experimental group spoke M = 8.27 syllables per pause group and M = 9.64 on the post-test (control group pre-test M = 10.47 and post-test M = 12.03). On the free-response task, changes from the pre- to post-test were flat for both groups. On the pre-test, the experimental group encoded short pause groups on average M = 5.55 syllables per pause group and M = 5.45 on the post-test (control group pre-test M = 5.56 and post-test M = 5.96).

In terms of unfilled planning and hesitation pauses, there were positive changes for both groups on the read-aloud task, but negative changes on the free-response task. On the read-aloud pre-test, the experimental group used M = 10.68 planning pauses per minute and M = 3.55 hesitation pauses per minute, meaning that 25% of all pauses were hesitation pauses. On the post-test this improved where the experimental group used M = 14.09 planning pauses per minute but only 2.48 hesitation pauses, meaning that only 15% of all pauses were hesitation pauses. On the read-aloud pre-test, the control group used M = 11.94 planning pauses per minute and M = 9.27 hesitation pauses (44% of all pauses were hesitation pauses), but on the post-test they used M = 15.63 planning pauses and only M = 3.00 hesitation pauses (only 15% were hesitation pauses). Note, however, the wide variation (large SD) on hesitation pauses for all participants on the read-aloud tasks, suggesting variable, idiosyncratic performances on this measure. It would be hard to claim these changes in either group were systematic on this measure.

On the free-response pre-test, the experimental group used only M = 4.55 planning pauses per minute, while using M = 11.08 hesitation pauses (71% were hesitations), and on the post-test they used M = 3.90 planning pauses per minute and M = 14.02 hesitation pauses (78% were hesitations). The control group used M = 5.71 planning pauses on the pre-test and M = 20.53 hesitation pauses per minute. Despite the difference from the experimental group in raw numbers, however, the control group’s proportion of hesitation pauses was 78%, not different from the experimental group’s profile. On the post-test, the control group used M = 5.49 planning pauses per minute with M = 21.5 hesitation pauses (80% were hesitations).

#### *5.1.4. Summary on pause groups*

To sum up, the results on speech rate suggested that explicit (and implicit) instruction in Discourse Intonation (DI) did not negatively affect learners’ overall fluency (see Table 1). When it came to length of pause groups and unfilled planning versus hesitation pauses, the results suggested an effect for explicit knowledge of DI (which both groups had) in that both groups improved on the read-aloud task on these measures. Participants in both groups had sufficient attentional resources to apply what they knew about DI, to speak in longer phrases, and to pause at appropriate places. When it came to the free-response task, however, neither group could summon up proceduralized explicit, or implicit, knowledge.

### 5.2. RQ #2 on prominence

Descriptive statistics for prominent syllables and unjustified prominent syllables for the read-aloud and free-response tasks are given in Tables 6 and 7.

Table 6 Descriptive statistics for prominent syllables and unjustified prominent syllables on the read-aloud task

| | Experimental Group | | Control Group | |
|---|---|---|---|---|
| | Pre-test | Post-test | Pre-test | Post-test |
| Prominent syllables (per minute) | *M* = 21.10<br>*SD* = 9.95 | *M* = 33.44<br>*SD* = 12.78 | *M* = 54.47<br>*SD* = 14.3 | *M* = 41.57<br>*SD* = 16.39 |
| Unjustified syllables (per minute) | *M* = 4.74<br>*SD* = 3.66 | *M* = 7.99<br>*SD* = 4.28 | *M* = 18.0<br>*SD* = 11.39 | *M* = 9.67<br>*SD* = 8.55 |

Table 7 Descriptive statistics for prominent syllables and unjustified prominent syllables on the free-response task

| | Experimental Group | | Control Group | |
|---|---|---|---|---|
| | Pre-test | Post-test | Pre-test | Post-test |
| Prominent syllables (per minute) | *M* = 13.46<br>*SD* = 4.74 | *M* = 15.63<br>*SD* = 11.27 | *M* = 28.27<br>*SD* = 18.23 | *M* = 22.69<br>*SD* = 19.49 |
| Unjustified syllables (per minute) | *M* = 2.53<br>*SD* = .72 | *M* = 4.14<br>*SD* = 5.60 | *M* = 9.41<br>*SD* = 8.94 | *M* = 5.65<br>*SD* = 5.8 |

#### *5.2.1. Task effects on prominence*

On the read-aloud task both groups used more prominence overall, whether it was placed appropriately or not. On the free-response task, both groups used fewer prominent syllables overall, as predicted (see Tables 6 and 7), likely due to greater processing demands. The proportion of unjustified prominent syllables did not increase in the free-response task, there were simply fewer of them. For instance, on the pre-test read-aloud task the experimental group used M = 21.10 prominent syllables per minute and M = 4.74 justified prominent syllables, meaning that 22% of the prominent syllables were placed inappropriately. On the free-response pre-test, the experimental group used M = 13.46 prominent syllables and M = 2.53 unjustified prominent syllables per minute (19% of all prominent syllables were inappropriate). Both groups had wide variations on unjustified prominence on both tasks, on both the pre- and post-tests suggesting much idiosyncracy on this measure.

A comparison with the native English speakers again underscores significant effects of increased processing burden on the free-response task, but also illuminates just how much more systematic native-English speakers are in their use of prominence. On the read-aloud tasks, one native English speaker used M = 43.48 prominent syllables per minute across the texts, and M = .66 unjustified prominent syllables (2% of the prominent syllables were inappropriate). On the free-response task a second native speaker used M = 33.47 prominent syllables per minute and M = 1.34 unjustified prominent syllables (4% of prominent syllables were inappropriate).

### *5.2.2. Developmental changes in prominence*

As with speech rate (see Results above), the experimental group was different from the control group at the outset of the study on both the read-aloud and the free-response task. Because they talked less overall, they also used less prominence. On the pre-test read-aloud, the experimental group used M = 21.10 prominent syllables per minute and M = 4.74 unjustified syllables (22% of the prominent syllables were inappropriate). The control group simply used more prominent syllables per minute (M = 54.47), but also used proportionately more unjustified syllables per minute (M = 18.00) on the pre-test, meaning that 33% of the prominent syllables were inappropriate. In the post-test read-aloud, both groups improved, but differently. The experimental group used more prominent syllables per minute (up to M = 33.44) and used only slightly more unjustified syllables (M = 7.99, or 24% of all prominent syllables). The control group used fewer prominent syllables per minute on the post-test, down to M = 41.57 which on its face does not seem like an improvement, but their use of unjustified prominent syllables per minute went down within the group more to M = 9.67 (23% of all prominent syllables) meaning that while the control group used fewer prominent syllables on the read-aloud, at least some of them were making better choices. Both the experimental and control groups ended up making about the same proportion of appropriate choices on the read-aloud (76%) at the end of the study.

On the free-response task the experimental group showed less obvious improvement, while the control group showed some improvement, the same as they did with the read-aloud task. On the free-response pre-test the experimental group used M = 13.46 prominent syllables per minute and M = 4.74 unjustified prominent syllables (around 19% of all syllables--a little less than the 22% of the read-aloud task). On the post-test the experimental group was using more prominent syllables per minute (M = 15.63), but also a larger proportion of unjustified prominent syllables (M = 4.14), meaning that 26% of the prominent syllables were inappropriate. On the free-response pre-test the control group used M = 28.27 prominent syllables per minute and M = 9.41 unjustified syllables (33% of all prominent syllables were inappropriate). On the post-test, while they used fewer prominent syllables overall (M = 22.69), they used proportionately fewer unjustified syllables per minute (M = 5.65), meaning that only 25% of their prominent syllables were inappropriate -- about the same level as their read-aloud post-test performance of 23%.

### *5.2.3. Summary on prominence*

To summarize, there was no clear evidence that the experimental group benefited from the implicit knowledge-building treatments. Further, the experimental group seems to have responded to the explicit instruction differently than the control group, and in ways that did not suggest greater proceduralized explicit knowledge for either group. In the explicit instruction, both groups were encouraged to use more prominence, and were given basic rules, linked to meaning, on prominence use. The experimental group did use more prominence in both post-tests, particularly on the read-aloud. They were able to muster enough explicit knowledge to make some appropriate choices on the read-aloud post-test, but may have been overgeneralizing prominence use and made more inappropriate choices on the post-test free-response task as processing demands increased. The control group, on the other hand, reduced their use of prominent syllables over time, but ended up increasing the proportion of appropriate choices (around 75%) even as their processing burdens increased in the free-response task. In their short experience with explicit instruction fresh in their minds, they may have decided to use what rules of prominence they knew but had to reduce how much prominence they used overall to compensate for the amount of attentional resources needed to use appropriate prominence. Once again, however, the large SDs on the unjustified prominence measure on both read-aloud and free-response tasks makes it difficult to make any claims on this issue (Tables 6 and 7).

### 5.3. RQ #3 on tone choices

Descriptive statistics for tone choices and syllables per tone choice for the read-aloud and free-response tasks are given in Tables 8 and 9.

Table 8 Descriptive statistics for rising, level, and falling tones and syllables per rising, level, and falling tone pause groups on read-aloud task

| | Experimental Group | | Control Group | |
|---|---|---|---|---|
| | Pre-test | Post-test | Pre-test | Post-test |
| Rising tones | *M* = .93 | *M* = 1.69 | *M* = 1.02 | *M* = .67 |
| (per minute) | *SD* = 1.37 | *SD* = 1.73 | *SD* = 1.96 | *SD* = 1.18 |
| Level tones | *M* = 4.23 | *M* = 4.8 | *M* = 8.17 | *M* = 5.28 |
| (per minute) | *SD* = 2.5 | *SD* = 3.82 | *SD* = 4.95 | *SD* = 3.25 |
| Falling tones | *M* = 9.38 | *M* = 10.13 | *M* = 12.02 | *M* = 12.68 |
| (per minute) | *SD* = 4.57 | *SD* = 5.29 | *SD* = 3.49 | *SD* = 2.87 |
| Syllables per rising tone | *M* = 4.0 | *M* = 9.36 | *M* = 1.5 | *M* = 1.14 |
| pause group | *SD* = 5.13 | *SD* = 10.2 | *SD* = 2.6 | *SD* = 1.95 |
| Syllables per level tone | *M* = 7.2 | *M* = 7.3 | *M* = 7.67 | *M* = 9.59 |
| pause group | *SD* = 2.31 | *SD* = 3.2 | *SD* = 6.0 | *SD* = 7.18 |
| Syllables per falling | *M* = 8.86 | *M* = 9.71 | *M* = 11.23 | *M* = 12.88 |
| tone pause group | *SD* = 1.34 | *SD* = 2.65 | *SD* = 3.29 | *SD* = 1.64 |

Table 9 Descriptive statistics for rising, level, and falling tones and syllables per rising, level, and falling tone pause groups on free-response task

| | Experimental Group | | Control Group | |
|---|---|---|---|---|
| | Pre-test | Post-test | Pre-test | Post-test |
| Rising tones | *M* = 2.37 | *M* = 2.06 | *M* = 2.06 | *M* = 3.3 |
| (per minute) | *SD* = 1.45 | *SD* = 1.21 | *SD* = 1.27 | *SD* = 3.94 |
| Level tones | *M* = 8.55 | *M* = 11.41 | *M* = 17.49 | *M* = 17.54 |
| (per minute) | *SD* = 2.25 | *SD* = 3.51 | *SD* = 6.97 | *SD* = 6.07 |
| Falling tones | *M* = 4.47 | *M* = 4.76 | *M* = 8.1 | *M* = 5.68 |
| (per minute) | *SD* = 1.1 | *SD* = 1.85 | *SD* = 2.52 | *SD* = 2.3 |
| Syllables per rising tone | *M* = 7.9 | *M* = 5.01 | *M* = 6.3 | *M* = 3.75 |
| pause group | *SD* = 2.52 | *SD* = 3.29 | *SD* = 4.7 | *SD* = 4.02 |
| Syllables per level tone | *M* = 4.41 | *M* = 3.63 | *M* = 4.68 | *M* = 4.9 |
| pause group | *SD* = 1.48 | *SD* = 1.23 | *SD* = 1.86 | *SD* = 2.18 |
| Syllables per falling | *M* = 7.29 | *M* = 7.42 | *M* = 7.39 | *M* = 8.85 |
| tone pause group | *SD* = 2.46 | *SD* = 2.05 | *SD* = 2.52 | *SD* = 1.6 |

### *5.3.1. Task effects on tone choices*

Interestingly, participants did something of the reverse with tone choices than they did with pause groups and prominence. With the pause groups and prominence, participants tended to do "better" on the read-aloud task, suggesting they could use their explicit DI knowledge when they had sufficient attentional resources to do so. When it came to the tone choice measures, participants used a greater variety of rising, level, and falling tones on the free-response task, while encoding only slightly shorter utterances within rising, level, and falling tone pause groups (against predictions made in Table 3). On the pre-test read-aloud the experimental group used M = .93 rising tones per minute, M = 4.23 level tones, and M = 9.38 falling tones (thus rising tones were 6% of all tone choices, level tones 29%, falling tones 65%). On the pre-test free response the experimental group used more rising tones per minute at M = 2.37 (15% of all tone choices), M = 8.55 level tones (63%), and M = 4.47 falling tones (29%). In terms of syllables per tone pause group for the pre-test read-aloud, the experimental group used M = 4.0 syllables per rising tone pause group, M = 7.2 syllables per level tone pause group, and M = 8.86 syllables per falling tone pause group. On the pre-test free-response task the experimental group used M = 7.9 syllables per rising tone pause group, M = 4.41 for level tones, and M = 7.29 for falling tones.

It might be argued that reading an academic passage aloud (read-aloud task) and talking casually about one's research (free-response task) are different acts, both rhetorically and socially. One might expect a speaker to use a greater variety of rising, level, and falling tones when interacting with other speakers. However, when comparing the native English speaker data from the read-aloud task the speaker used M = 10.42 rising tones per minute across the four texts (or 50% of all tone choices used), only M = 2.56 level tones (or 12%), and M = 7.79 falling tones (or 38%). The native English speaker used far more rising tones and far fewer level tones when reading college-level science passages aloud. The native English speaker apparently used rising and falling tone choices as an expression of discoursal meaning differently than the participants in either group. The same general pattern followed for the free-response task. The second native speaker used M = 9.37 rising tones per minute (or 44% of all tone choices used), M = 8.03 level tones (37%), and only M = 4.02 falling tones (19%).

On the read-aloud task, one native English speaker used M = 10.43 syllables per rising tone pause group, M = 9.88 syllables per level tone pause group, and M = 10.92 syllables per falling tone pause group. On the free-response task, the second native speaker used even longer utterances on rising and falling tone choice pause groups with M = 14.57 syllables per rising tone pause group, M = 2.5 syllables for level tones, and M = 12.33 syllables for falling tones. The native English speakers encoded longer utterances on both tasks for rising and falling tone choice pause groups, with level tone choice pause groups getting the shortest utterances on average, in contrast to the non-native English-speaking experimental and control groups who tended to use longer utterances in level tone choice pause groups, when compared to both rising and falling tone choice pause groups.

### *5.3.2. Developmental changes in tone choices*

The experimental group showed somewhat more improvement than the control group on the read-aloud, if one counts improvement as participants using a higher proportion of rising tones, and encoding longer rising tones. On the pre-test read-aloud the experimental group used M = .93 rising tones per minute (6% of all tone choices), M = 4.23 level tones

(29%), and M = 9.38 falling tones (65%). On the post-test read-aloud they used M = 1.69 rising tones per minute (10% of all tone choices), M = 4.8 level tones (29%), and M = 10.13 falling tones (61%). Some of the experimental groups were also able to encode longer rising tone choice pause groups on the pre-test, suggesting a discoursal knowledge of rising tones. On the pre-test read-aloud, they used M = 4.0 syllables per rising tone, but used M = 9.36 on the post-test. On the pre-test, they used M = 7.2 syllables per level tone, and M = 8.86 syllables per falling tone. For the post-test the experimental group used M = 7.3 syllables per level tone and M = 9.71 syllables per falling tone.

The control group did not seem to improve on the read-aloud task. On the pre-test they used M = 1.02 rising tones per minute (5% of all tone choices used), M = 8.17 level tones (39%), and M = 12.02 falling tones (57%). On the post-test they only used M = .67 rising tones (4%), M = 5.28 level tones (39%), and M = 12.68 falling tones (68%). Their rising tone choice pause groups were very short, widely variable (large SDs), and did not change over time. On the pre-test read-aloud the control group used M = 1.5 syllables per pause group, and on the post-test, they used only M = 1.14. The length of their level and falling tones remained about the same over time. On the pre-test they used M = 7.67 syllables per level tone and M = 11.23 per falling tone. On the post-test they used M = 9.59 syllables per level tone and M = 12.88 syllables per falling tone.

On the free-response task, there was little evidence of strong change on the part of either group. On the free-response pre-test, the experimental group used M = 2.37 rising tones per minute (15%), M = 8.55 level tones (56%), and M = 4.47 falling tones (29%). On the post-test they used M = 2.06 rising tones per minute (down to 11%), M = 11.41 level tones (up to 63%), and M = 4.76 falling tones (26%). Perhaps the experimental group used level tone choices discoursally in place of rising tone choices. On the free-response pre-test, the control group used M = 2.06 rising tones per minute (7% of all tone choices), M = 17.49 level tones (63%), and M = 8.1 falling tones (29%). On the post-test, their use of rising tones increased a little to M = 3.3 per minute (12%)(note, however the large SD which suggested extreme within-group variability). Their use of level and falling tones on the post-test was M = 17.49 level tones per minute (63%), with slightly fewer falling tones at M = 5.68 (21%).

While the experimental group consistently encoded longer rising tone choice pause groups than the control group, there was little evidence of improvement. On the pre-test, the experimental group encoded M = 7.9 syllables per pause group, while the control group encoded only M = 6.3. Both groups did somewhat worse on the post-test with the experimental group using M = 5.01 syllables per pause group and the control group using only M= 3.75. Length of level and falling tone choice pause groups remained unchanged for both groups from the pre- to the post-test free-response task.

#### *5.3.3. Summary on tone choices*

The experimental group showed more improvement on the read-aloud task than the control group, and was able to use a greater proportion of rising tones which some were able to encode in longer utterances. This suggested very proceduralized knowledge (or perhaps implicit knowledge) that they could apply to the read-aloud task. There was little evidence of improvement on the free-response task for either group, although there was a tendency for the experimental group to encode longer rising tone choice pause groups than the control group. Even though the experimental group declined on this measure over time, it still suggested perhaps implicit knowledge on their part that the three tone choices were discoursally different to them. The longer encodings that some experimental group members achieved suggested they wished to say something substantive using a rising tone choice. Whether this came from their experience using English in the U.S. or from explicit or implicit instruction is not known.

## 6. Discussion

### 6.1. Good news

The good news was that explicit and implicit instruction in DI did not slow down participants' speech rates, suggesting that learners did not have to divert attentional resources to focus on form (DI). Further, explicit knowledge made (in most cases) a positive difference on the various measures of DI, on the read-aloud task. Many of the predictions of developmental changes made in Tables, 1, 2, and 3 were confirmed insofar as they concerned the read-aloud task and explicit knowledge. Finally, the read-aloud task, while not approximating authentic language use, did reveal a good deal about both groups' DI use. Their performances on both the pre- and post-tests might be taken to represent an idealized picture of their current state of DI development. Task types do strongly influence learners' performance of DI (see predictions in Tables 1, 2, and 3, and the Results), particularly with regards to attentional load, and this needs to be taken into account when assessing DI. It is not that one type of task is better than the other. Rather, they tell us different things about the type of knowledge learners are using.

### 6.2. Bad news

There were few observable changes in the experimental group's implicit knowledge or use of DI, as evidenced across the board by the free-response task. As a colleague put it, "acquiring L2 implicit knowledge is really just a hard, slow slog through a muddy field." One might argue that in actual teaching, ITAs may take greater than ordinary care in speaking (comprehensible output) and do better with their DI, if they realize it is useful for expressing intended meanings in classrooms. But whether or not they would draw on implicit knowledge of DI to do this is not known. In fact, in one case it was shown that when an ITA had been approved to teach, she stopped using prominence in her teaching talk (Lee and Gorsuch 2012), suggesting that she used explicit knowledge to pass her ITA course teaching simulation requirements. Still, it was shown in this study that a free-response task was an effective way to "get at" changes in implicit knowledge, for better or worse.

It did not help that the two groups seemed different on the measures to begin with and it raised doubt as to whether the comparison was fair. Yet the "stalled out" experimental group with one semester in the U.S. and the newly arrived control group with no experience in an English-speaking country, but with more variable English ability, represented two real populations in ITA programs. These are the populations we have to work with, and the comparison did show a few DI features which may represent differences in implicit knowledge due to experience using English on campuses in the U.S. First, the experimental group did show that they used more rising tones in a general sense, even at the beginning of the study. Their speech was slightly more "musical" with a variety of rising and falling tones. Second, and more importantly, some members of the experimental group encoded longer rising tone pause groups in both read-aloud and free-response tasks, suggesting they had some implicit knowledge of the discoursal value of rising tones.

### 6.3. Limitations and suggestions for future research

The number of participants was small, and as with other fluency development research, the results highly variable, possibly because of learners' idiosyncratic development (see Freed 1995). Further, only Mandarin L1 speakers were studied. Explicit and implicit knowledge growth may be different with other L1 groups. The measures themselves need further study, to establish better precision in measuring speaking fluency development. Qualitative accounts of listener perception could also be taken into account. For instance, listeners may have judged the experimental group more comprehensible because they were speaking more slowly. A qualitative analysis of the transcripts themselves would also aid in learning to know more about the difference, if any, of the nature of the use of prominence (both appropriate and inappropriate) between groups.

While implicit knowledge growth is slow, this is not a reason to disregard the need for finding ways to develop learners' implicit knowledge of DI. This is especially true if it can be done in pedagogically worthwhile ways. The audio-supported Repeated Reading treatments arguably are pedagogically worthwhile, as they support vocabulary building (needed for ITAs to know where to use prominence; see for example Wennerstrom 1998) and provided consistent models of DI connected to discoursal meaning. One question, however, is whether the texts for the RR were too long. In other words, did the experimental group simply have too many things in the visual and aural input to pay attention to? Further, did the experimental group actually hear the DI features in the audio support parts of RR? A new experiment should be done to explore this issue focusing on the idea of, if the participants cannot demonstrate awareness of DI features immediately after hearing them (say, by repeating the last heard sentence of text aloud), can they be said to have noticed the feature? This may also show whether some aspects of DI, particularly prominence, are simply harder to detect in aural input.

## 7. Conclusion

This study explored a theoretically-driven permutation of an intervention designed to improve ITAs' spoken DI. Of particular interest was to learn, if effects of implicit knowledge growth in DI could be found as the result of an experimental group participating in audio-assisted repeated reading treatments using twice-weekly easy, popular science texts for 14 weeks. In a read-aloud condition, where speech processing burdens were reduced, both the experimental and control groups improved over time on speech rate, planning pauses versus hesitation pauses, prominence, tone choices, and length of tone choices. This suggested participants had sufficient attentional resources, and explicit knowledge of DI, to apply DI to their oral renderings. In a free-response task, processing burdens were increased as participants had to encode their own speech. On most measures there was little evidence of change in implicit knowledge of DI. Clearly, implicit knowledge is difficult to form and change. At the same time, explicit DI instruction did not reduce participants' overall fluency measure, that of speech rate. Explicit DI instruction where form is linked to meaning is worthwhile, in that explicit knowledge may become proceduralized and available for learners' extemporaneous use.

Implicit knowledge building in DI, while difficult to demonstrate, may still be worthwhile, if it builds learners' knowledge of vocabulary (to improve prominence, or sentence level stress) and builds their experience hearing DI features linked to meaning within extended texts.

## REFERENCES


Abe, H. The effect of interactive input enhancement on the acquisition of the English connected speech by Japanese college students. *Phonetics Teaching & Learning Proceedings, University College London*. Retrieved June 19, 2013, from http://www.phon.ucl.ac.uk/ptlc/proceedings/ptlcpaper_29e.pdf. 2009.

Anderson-Hsieh, J. & Koehler, K. The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38,* 561-570. 1988.

Arevart, S. & Nation, P. Fluency improvement in a second language. *RELC Journal, 22,* 84-94. 1991.

Bolinger, D. Intonation in American English. In D. Hirst & A. Di Cristo (Eds.). *Intonation systems: A survey of twenty languages* (pp. 45-55). Cambridge: Cambridge University Press. 1998.

Brazil, D. *The communicative value of intonation in English*. Cambridge: Cambridge University Press. 1997.

Butcher, A. Pause and syntactic structure. In H. Dechert & M. Raupach (Eds.). *Temporal variables in speech* (pp. 85-90). The Hague: Mouton Publishers. 2001.

Butterworth, B. Evidence from pauses in speech. In B. Butterworth (Ed.). *Language production: Speech and talk* (pp. 155-176). London: Academic Press. 1980.

Cardillo, M. You know you are at UConn when...[Letter to the editor]. *Daily Campus,* p. 4. 2002, February 21.

Chambers, F. What do we mean by fluency? *System, 25*(4), 535-544. 1997.

Chiang, S. Dealing with communication problems in the instructional interactions between international teaching assistants and American college students. *Language and Education, 23*(5), 461-478. 2009.

Chun, D., Hardison, D., & Pennington, M. Technologies for prosody in context. In J. Hansen Edwards & M. Zampini (Eds.). *Phonology and second language acquisition*. (pp. 323-346). Amsterdam: John Benjamins Publishing Company. 2008.

Council of Graduate Schools *Findings from the 2010 CGS international graduate admissions survey phase III: Final offers of admission and enrollment November 2010*. Washington, D.C.: Author. 2010.

Crystal, D. & Davy, D. *Investigating English style*. Bloomington, IN: Indiana University Press. 1969.

Dechert, H. Second language production: Six hypotheses. In H. Dechert, D. Mohle, & M. Raupach (Eds.). *Second language productions* (pp. 211-230). Tubingen, Germany: Gunter Narr Verlag Tubingen. 1984.

Deese, J. Pauses, prosody, and the demands of production in language. In H. Dechert & M. Raupach (Eds.). *Temporal variables in speech* (pp. 69-84). The Hague: Mouton Publishers. 1980.

Doughty, C. Instructed SLA: Constraints, compensation, and enhancement. In C. Doughty & M. Long (Eds.). *The handbook of second language acquisition* (pp. 256-310). Malden, MA: Blackwell Publishing Ltd. 2003.

Educational Testing Service (1996). *The SPEAK rater training kit*. Princeton, NJ: Author.

Ejzenberg, R. The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.). *Perspectives on fluency* (pp. 287-313). Ann Arbor, MI: The University of Michigan Press. 2000.

Ellis, R. *SLA research and language teaching.* Oxford: Oxford University Press. 1997.

Escudero, P. Second-language phonology: The role of perception. In M. Pennington (Ed.). *Phonology in context* (pp. 109-134). Houndmills, United Kingdom: Palgrave MacMillan. 2007.

Field, J. *Listening in the language classroom*. Cambridge: Cambridge University Press. 2000.

Field, J. Bricks or mortar: Which parts of the input does a second language listener rely on? *TESOL Quarterly, 42*(3), 411-432. 2008.

Ford, J., Gappa, L., Wendorff, J. & Wright, D. Model of an ITA institute. In J. Nyquist, R. Abbott, D. Wulff & J. Sprague (Eds.). *Preparing the professoriate of tomorrow to teach* (pp. 341-349). Dubuque, IA: Kendall Hunt. 1991.

Foster, P., Tonkyn, A. & Wigglesworth, G. Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354-375. 2000.

Freed, B. What makes us think that students who study abroad become fluent? In B. Freed (Ed.). *Second language acquisition in a study abroad context* (pp. 123-148). Amsterdam: John Benjamins Publishing Company. 1995.

Gibson, S. Reading aloud: a useful learning tool? *ELT Journal, 62*(1), 29-36. 2008.

Gilbert, J. Pronunciation practice as an aid to listening comprehension. In D. Mendelsohn & J. Rubin (Eds.). *A guide for the teaching of second language listening* (pp. 97-112). San Diego, CA: Dominie Press, Inc. 1995.

Goldman Eisler, F. *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press. 1968.

Gorsuch, G.J. "Exporting English pronunciation from China: The communication needs of young Chinese scientists as teachers in higher education abroad." *Forum on Public Policy*, 3. 2011a. Available: http://forumonpublicpolicy.com/vol2011no3/archive/gorsuch.pdf

Gorsuch, G. J. "Improving speaking fluency for International Teaching Assistants by increasing input." *TESL-EJ*, 14(4), 1-25. Available: http://www.tesl-ej.org/wordpress/issues/volume14/ej56/ej56a1/ 2011b.

Gorsuch, G., Meyers, C., Pickering, L., and Griffee, D.T. *English Communication for International Teaching Assistants*. Long Grove, IL: Waveland Press, Inc. 2010.

Gorsuch, G., Meyers, C., Pickering, L., and Griffee, D.T. *English Communication for International Teaching Assistants* (2nd ed.). Long Grove, IL: Waveland Press, Inc. 2013.

Gorsuch, G. and Sokolowski, J. "International teaching assistants and summative and formative student evaluation." *Journal of Faculty Development*, 21(2): 117-136. 2007.

Gorsuch, G.J. and Taguchi, E. "Repeated reading for developing reading fluency and reading comprehension: The case of EFL learners in Vietnam." *System*, 36(2): 253-278. 2008.

Gorsuch, G.J. and Taguchi, E. "Developing reading fluency and comprehension using repeated reading: Evidence from longitudinal student reports." *Language Teaching Research,* 14(1): 27-59. 2010.

Griffee, D.T., Gorsuch, G., Britton, D. and Clardy, C. "Intensive second language instruction for international teaching assistants: How much is effective?" In *Teachers and Teaching: Strategies, Innovations and Problem Solving*, edited by G. Ollington, 187-205. Hauppage, NY: Nova Science Publishers, Inc. 2009.

Hahn, L. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*(2), 201-223. 2004.

Hirst, D. & Di Cristo, A. *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press. 1998.

Kaufman, D. & Brownworth, B. Collaborative paradigms and future directions in international teaching assistant professional development. In D. Kaufman & B. Brownworth (Eds.). *Professional development of international teaching assistants* (pp. 1-13). Alexandria, VA: Teachers of English to Speakers of Other Languages. 2006.

Kiany, G.R. & Shiramiry, E. The effect of frequent dictation on the listening comprehension ability of elementary EFL learners. *TESL Canada Journal, 20*(1), 57-63. 2002.

Kormos, J. & Demes, M. Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*, 145-164. 2004.

Lee, E. & Gorsuch, G. Conversational Analysis of the classroom communication of a math International Teaching Assistant. In G. Gorsuch (Ed.). *Working theories for teaching assistant development* (pp. 661-689). Stillwater, OK: New Forums Press. 2012.

Levis, J., Muller Levis, G. & Slater, T. Written English into spoken: A functional discourse analysis of American, Indian, and Chinese TA presentations. In *Working theories for teaching assistant development* (pp. 521- 565). Stillwater, OK: New Forums Press. 2012.

Lennon, P. The lexical element in spoken second language fluency. In H. Riggenbach (Ed.). *Perspectives on fluency* (pp. 25-42). Ann Arbor, MI: The University of Michigan Press. 2000.

Levelt, W. *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press. 1989.

McKeachie, W. *Teaching tips* (11th ed.). Boston: Houghton Mifflin Company. 2004.

Munro, M. Foreign accent and speech intelligibility. In J. Hansen Edwards & M. Zampini (Eds.). *Phonology and second language acquisition* (pp. 193-218). Amsterdam: John Benjamins Publishing Company. 2008.

Olynak, M., Anglejan, A., & Sankoff, D. A quantitative and qualitative analysis of speech markers in the native and second language speech of bilinguals. In R. Scarcella, E. Andersen, & S. Krashen (Eds.). *Developing communicative competence in a second language* (pp. 139-155). Boston, MA: Heinle & Heinle Publishers. 1990.

Pawley, A. & Syder, F. The one-clause-at-a-time hypothesis. In H. Riggenbach (Ed.). *Perspectives on fluency* (pp. 163-199). Ann Arbor, MI: The University of Michigan Press. 2000.

Pennington, M. *Phonology in English language teaching*. London: Longman. 1996.

Pennington, M. & Ellis, N. Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal*, *84*(3), 372-389. 2000.

Pickering, L. *An analysis of prosodic systems in the classroom discourse of native speaker and nonnative speaker teaching assistants*. Unpublished doctoral dissertation. University of Florida, Gainesville, FL. 1999.

Pickering, L. The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, *35*(2), 233-255. 2001.

Pickering, L. Intonation. In K. Malmkær (Ed.). *The Routledge linguistics encyclopedia* (3rd ed.)(pp. 280-286). London: Routledge. 2010.

Riggenbach, H. *Nonnative fluency in dialogue versus monologue speech: A microanalytic approach*. Unpublished doctoral dissertation. University of California at Los Angeles. 1989.

Schmidt, R. Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14,* 357-385. 1992.

Skehan, P. Task-based instruction. *Language Teaching, 36,* 1-14. 2003.

Smith, R., Byrd, P., Nelson, G., Barrett, R. & Constantinides, J. *Crossing pedagogical oceans: International teaching assistants in U.S. undergraduate education*. Washington, D.C.: ERIC Clearinghouse on Higher Education. 1992.

Strange, W. & Shafer, V. Speech perception in second language learners: The re-education of selective perception. In J. Hansen Edwards & M. Zampini (Eds.). *Phonology and second language acquisition* (pp. 153-191). Amsterdam: John Benjamins Publishing Company. 2008.

Tavakoli, P. Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal, 65*(1), 71-79. 2010.

Tench, P. *The intonation system of English*. London: Cassell. 1996.

Wennerstrom, A. Intonation and second language acquisition: A study of Chinese speakers of English. *Studies in Second Language Acquisition, 20*(1), 1-25. 1998.

Wennerstrom, A. The role of intonation in second language fluency. In H. Riggenbach (Ed.). *Perspectives on fluency* (pp. 102-127). Ann Arbor, MI: The University of Michigan Press. 2000.

Wennerstrom, A. *The music of everyday speech*. Oxford: Oxford University Press. 2001.

Williams, J. Planning, discourse marking, and the comprehensibility of international teaching assistants. *TESOL Quarterly*, *26*, 693-711. 1992.

Young, L. *Crosstalk and culture in Sino-American communication*. Cambridge: Cambridge University Press. 1994.

Zampini, M. L2 speech production research. In J. Hansen Edwards & M. Zampini (Eds.). *Phonology and second language acquisition* (pp. 219-249). Amsterdam: John Benjamins Publishing Company. 2008.