Running Head: Phonology, Mixed Methods

Phonology, Mixed Methods

Talia Isaacs, University of Bristol

Corresponding author: Talia Isaacs

Graduate School of Education

University of Bristol

35 Berkeley Square, Clifton

Bristol, United Kingdom BS8 2AH

talia.isaacs@bristol.ac.uk

Isaacs, T. (2013). Phonology: Mixed methods. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (Vol. 7, pp. 4427-4434). Hoboken, NJ: Wiley-Blackwell.

[A]Background

A growing number of researchers in the area of second language **phonology** have employed a **mixed methods** approach to better understand the research problem and to strengthen the quality of their inferences or interpretations (e.g., Derwing & Munro, 2009; Zielinski, 2008). However, not all published mixed methods phonological studies are explicitly labeled as "mixed methods," which may make them difficult to locate in tables of contents or on-line searches. Furthermore, the authors may not emphasize their reasons for using more than one method or describe the research design in much detail. The following is an example of a mixed methods study situated within the subdiscipline of second language (L2) phonological research. The focus is to address the rationale for integrating **quantitative** and **qualitative** methods, to discuss the nature of the mixing that occurred in the data collection and analyses, and to emphasize how quantitative and qualitative data were converged to yield rich empirical outcomes relevant to the research purpose.

The study arose from the widely-held view among pronunciation experts that **intelligibility** should be the foremost goal of L2 **pronunciation instruction** (Morley, 1991; Munro & Derwing, 1999). However, it remained to be explored whether intelligibility is an appropriate criterion for assessing the pronunciation of nonnative speaking graduate students in academic settings. Many **international teaching assistants** (ITAs), for example, are expected to carry out instructional duties in their L2 in addition to their academic tasks. While several factors are clearly important for any instructor's pedagogical success, pronunciation is often identified as the culprit for poor ITA performance from the perspective of language experts, students, and ITAs themselves (Cheng, Myles, & Curtis, 2004; Hoekje & Williams, 1992). Situated at a North American university where ITA screening does not currently take place, the present study sought to examine whether the criterion of intelligibility is "enough" (i.e., a sufficient goal) for assessing pronunciation proficiency in nonnative English speaking graduate students.

A secondary goal was to explore which aspects of pronunciation most interfere with intelligibility from the perspective of both undergraduate students, who have a stake in understanding ITAs in real-world contexts, and a pronunciation expert (the researcher).

Ultimately, it was the need to adequately address the research questions that dictated the decision to mix methods. The reasoning was that drawing on a single method or source of evidence would have been too restrictive in shedding light on the complex phenomenon of intelligibility, which encompasses both the speech **production** and **perception** (Morley, 1991), and for evaluating its suitability as an **assessment** criterion for nonnative graduate students. This philosophical orientation of doing "what works" in order to answer the research questions, often referred to as **pragmatism** in the mixed methods literature, was embraced in this study (Greene, 2007, pp. 83–85; Teddlie & Tashakkori, 2009).

[Data Collection Procedure]

For the purpose of this investigation, intelligibility was defined as the percent of a nonnative speaker's words that listeners indicate being able to understand. Speech samples of eight nonnative English speaking graduate students telling a story and expressing an opinion were audio recorded, randomized, and burned onto CD. Eighteen native English speaking undergraduate students then listened to and assessed each speaker's intelligibility on a 0–100% **rating scale**. After a second listening, the listeners consulted a list of potential pronunciation problem areas (e.g., individual sounds, word stress, sentence rhythm, speech rate, pitch, and speech clarity) and rank ordered the top three features, if any, that had most hindered their understanding of the speaker's words. Within each ranked category, they were then asked to identify the source of the problem by selecting one of two options. For example, if they had selected "speech clarity," they were asked to identify whether the most prominent problem was that the speaker "overpronounces" or "mumbles" (see Isaacs, 2008b, for more on procedures and instruments). Finally, the listeners indicated whether they felt that the speaker's

pronunciation was sufficient for him/her to instruct an undergraduate course as a teaching assistant (TA).

The listeners' responses described above, which were a quantitative source of evidence, were obtained through closed-ended questionnaire items. In fact, the same listener questionnaire was used to collect qualitative data simultaneously through open-ended questions. The listeners were invited to provide written comments on their impressions of the speech at any time during the rating session. *Figure 1* illustrates the research design of this study, which is classified as a mixed methods **triangulation** design in Creswell and Plano Clark's typology (2007, pp. 62–67). The interplay between quantitative and qualitative methods, where one method is used to clarify or expand upon findings from the other, is considered an advantage of this design.

< INSERT FIGURE 1 ABOUT HERE>

[A]Data Analysis

The qualitative speech analysis was a data-driven process that involved the **transcription** of the speech data into **phonetic** symbols, the development of a transcription system for **suprasegmentals**, and the use of **color-coding** to identify unintelligible words. The researcher color-coded the data twice during separate listenings to check for consistency. Then, the qualitative data were "quantitized" (Teddlie & Tashakkori, 2009, p. 343), or transformed into quantitative data, by counting the total number of coded instances of unintelligibility, so that the ratio of unintelligible words over the total number of words could be calculated for each speaker. The speech samples were additionally analyzed quantitatively for speaking rate, or the number of syllables uttered per second, and articulation rate, or the number of syllables uttered per second excluding pauses that exceeded .5 seconds (Munro & Derwing, 1998, pp. 166-167).

< INSERT FIGURE 2 ABOUT HERE>

< INSERT FIGURE 3 ABOUT HERE>

The quantitative listener data were analyzed using descriptive statistics and frequencies in *SPSS 17.0*. Cronbach's alpha was computed preliminarily as a measure of **interrater reliability** by treating the 18 sets of intelligibility ratings as item scores. In preparation for analysis, the spreadsheet was organized by listing the eight L2 speakers in separate rows and the 18 raters in separate columns. An alpha coefficient of .883 revealed that raters were overall consistent in their judgments.

< INSERT FIGURE 4 ABOUT HERE>

The qualitative listener comments were analyzed iteratively using an adaptation of Strauss and Corbin's **open-coding**, with the overall goal of generating categories from the data (1990). First, the data were transcribed in a word processor using speaker and listener identification codes. Next, the comments were grouped together according to a perceived common theme or idea, and meaning-laden words were bolded (e.g., "**trails off** on sentences. Doesn't **finish** her **thoughts**" was grouped with "**thoughts** seem **separated** & **disconnected**"). The listeners' own language was then used to label the categories and subcategories. Finally, the researcher's language was imposed on the category labels where necessary to clarify meaning (e.g., the subcategory previously labeled, "Pronunciation is irritating but does not affect clarity of what is said" was more precisely designated, "A feature of the speaker's pronunciation, while noticeable or irritating, does not affect overall intelligibility"). In order to check the consistency of the categorizations, the researcher randomized and recoded the data a few weeks later. Similar categories emerged both times, which revealed a high level of intracoder reliability (Johnson & Christensen, 2008).

The quantitative and qualitative strands from the analysis of the speech and listener data were merged in an attempt to effectively answer the research questions. Specific cases of a few speakers will be illuminated in presentation of the results in the manner of a **case study** (see Yin, 2003).

[A]Results

The first research question examined whether intelligibility is a sufficient criterion for assessing the pronunciation of nonnative speaking English graduate students. Table 1 shows that Speaker K emerged as the most intelligible speaker overall. The percent of his words that the listeners indicated being able to understand was about 9% higher than that of Speaker F, the second highest performing speaker. In addition, Speaker K's low standard deviation relative to all other speakers indicated that the listeners reached the greatest consensus about his degree of intelligibility. At the other extreme, Speaker C was regarded as the least intelligible speaker. Her mean intelligibility rating was over 20% lower than that of the second least intelligible speaker (Speaker M) and was nearly 30% below the group mean.

Table 1

Mean intelligibility ratings grouped by speaker

Intelligibility ratings			
Speaker	Mean (%)	*SD*	Ranking
C	46.9	18.5	8
E	75.2	15.1	6
F	86.3	13.7	2
G	76.6	14.0	5
K	95.2	8.5	1
M	67.7	18.2	7
N	79.1	18.2	4
R	85.8	12.6	3
Group	76.6	20.2	

Note. Means percentages are based on 18 ratings for each speaker and 144 ratings for the whole group.

Figure 5 shows that no listeners felt that Speaker C's pronunciation was adequate for her to instruct an undergraduate course as a TA (hereafter referred to as the "TA question"), although 1 in 6 were uncertain. Speaker M, the speaker with the second lowest intelligibility rating, was the only other speaker who received more negative than positive responses to the TA question, although her outcome

was better than Speaker C's. At the other end of the spectrum, Speaker K's top ranking for intelligibility was not maintained on the TA question. Clearly, a majority of listeners endorsed Speaker K as a TA. However, Speaker F surpassed him as the top ranked speaker by receiving slightly more positive responses and fewer negatives. Although this difference between Speaker K's and Speaker F's scores may seem negligible, it was key to understanding whether intelligibility is "enough" for ITAs from the perspective of undergraduate students. In fact, two listeners who had rated Speaker K's words as 100% intelligible responded "no" to the TA question. One elaborated, "only a few problematic words, but would be quite annoying as a TA," and the other commented, "far too slow, as a TA, he would be quite boring" (Isaacs, 2008a, p. 52). Additional evidence was needed to fully probe why Speaker K's nearly fully intelligible pronunciation did not translate into the same degree of consensus among undergraduate listeners about the adequacy of his pronunciation to serve as a TA with instructional duties.

<INSERT FIGURE 5 ABOUT HERE>

The researcher's quantitative analysis of Speaker K's fluency showed that his speaking rate was the slowest overall (see Figure 6). In contrast, his mean articulation rate was over 1 syllable/second faster than that of the second fastest speaker (Speaker N). The suprasegmental transcriptions, which included the use of textual symbols for pauses and tempo, indicated that Speaker K consistently paused at the end of thought groups. So while Speaker K spoke relatively quickly during his phonation time (articulation rate) and paused at logical junctures, his pauses accounted for a substantial portion (62%) of his recorded speech. This can at least partially account for the following listener comments:

- I feel frustrated waiting for him to get on with what he is trying to say.
- Though slow, the choice of words is excellent and the effect is soothing.
- Would be very good if he sped up to create better flow.
- Hard to link parts of sentence together.

< INSERT FIGURE 6 ABOUT HERE>

Some listeners additionally remarked on Speaker K's monotone speech. The overall effect of Speaker K's speech and the presiding negative sentiment about his long pauses are perhaps best summarized by the following listener comment: "I understood all he said, but he spoke slowly and had no intonation in his voice which made it quite obnoxious." This implies that while Speaker K's words were largely intelligible, intelligibility was, at least in some listeners' views, an insufficient condition for him to instruct an undergraduate course as a TA. Nonetheless, intelligibility appeared to be a pre-requisite for serving as a TA from the listeners' (undergraduate students') perspective. There was absolute consensus, for example, that Speaker C, who was judged to be the least intelligible speaker overall, should not be charged with a TA position on the basis of her pronunciation. It may be that there is a threshold level of intelligibility that is necessary to carry out instructional responsibilities as a TA for an undergraduate course. In this study, it appears that Speaker C is decidedly below that threshold. Her *z*-score (standard score) for intelligibility was the only one in the data set that exceeded one standard deviation below the mean ($z = -2.03$), although a precise cut-off was not defined.

Table 2 shows the pronunciation features that the listeners identified as being the most problematic. The caveat is that they did not appear to be able to distinguish the features that rendered a speaker unintelligible from those that, while noticeable or irritating, did impede intelligibility. The problem areas with the highest listener response frequencies were associated with Speaker C. From the standpoint of intelligibility based on the researcher's coding, what was most problematic about Speaker C's speech was, in the listeners' words, that "she sort of stopped & started," "rhythm very broken," "stutters, hesitates," "lots of unsure pauses" and "words are slurred incomprehensibly" (Isaacs, 2008a, p. 60).

Table 2

Percent of raters who identified the pronunciation feature as being the "most prominent problem" (raw frequency) grouped by speaker.

Speech category	Problem cited by listeners	Speakers							
		C	E	F	G	K	M	N	R
Segmentals	Substitute sounds	72% (10)	31% (5)	22% (4)	41% (7)	-	41% (7)	-	50% (9)
	Delete/add sounds	35% (6)	53% (9)	-	35% (6)	-	47% (8)	28% (5)	-
Word stress	Wrong stress syll.	38% (7)	-	44% (8)	56% (10)	-	50% (9)	-	33% (6)
	Distinguish syllables	-	-	-	-	-	28% (5)	-	-
Sentence Rhythm	Distinguish words	33% (6)	39% (7)	44% (8)	33% (6)	-	28% (5)	-	28% (5)
	Linking	-	22% (4)	-	-	-	-	-	-
Rate of Speech	Too fast	-	-	-	-	-	-	50% (9)	-
	Too slow	-	-	-	22% (4)	67% (12)	-	-	-
Pitch	Pitch change	-	-	22% (4)	-	-	-	-	-
	Monotone	-	-	-	-	44% (8)	-	-	-
Speech Clarify	Overpronunces	-	-	28% (5)	-	-	24% (4)	-	-
	Mumbles	72% (13)	22% (4)	-	28% (5)	-	41% (7)	67% (12)	29% (6)

Note. Only pronunciation features identified by 20% or more of raters are shown. Missing data are excluded from the percentage calculations.

Speaker C's abrupt utterance [»n̩ wɛn aj« (1.) wɛn aj] shown in *Figure 7* remained unintelligible to the researcher even after multiple listenings at a reduced speed. In contrast to Speaker K, Speaker C's pauses did not appear at the end of thought groups, which may have led some listeners to perceive them as sounding "awkward." The jarring "slow to fast to stop to start" of her "erratic speech," as some undergraduate listeners described, was analogized as sounding like the sputtering engine of a stalled car by the researcher in a research memo.

[A]Concluding Remarks

Taken together, the listeners' and researcher's analysis of the speech did not provide conclusive evidence as to which pronunciation features most impede intelligibility. However, the different sources of evidence appeared to reveal that the problem areas identified under the categories "individual sounds," "word stress," and "speech clarity" played the most important role in the researchers' and undergraduate listeners' perceptions of unintelligibility. Conversely, "sentence rhythm," "rate of speech," and "pitch" only tended to impede perceptions of intelligibility when they acted in conjunction with another problematic pronunciation feature. More research is needed to isolate pronunciation features and examine impacts on intelligibility (e.g., Hahn, 2004). Such research is likely to be enriched by the different perspectives offered by multiple methods, where the strengths of one method compensate for the weaknesses of the other.

< INSERT FIGURE 7 ABOUT HERE>

SEE ALSO: Intelligibility; Mixed Methods; Phonetics and Phonology; Pronunciation Assessment; Rating Oral Language; Speech Perception; Transcription

References

Cheng, L., Myles, J., & Curtis, A. (2004). Targeting language support for non-native English-speaking graduate students at a Canadian university. *TESL Canada Journal, 21*(2), 50-71.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.

Derwing, T.M. & Munro, M.J. (2009). Comprehensibility as a factor in listener interaction preferences: Implications for the workplace. *Canadian Modern Language Review, 66*(2), 181-202.

Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: Jossey-Bass.

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*(2), 201-233.

Hoekje, B., & Williams, J. (1992). Communicative competence and the dilemma of international teaching assistant education. *TESOL Quarterly, 26*(2), 243–269.

Isaacs, T. (2008a). *Assessing second language pronunciation: A mixed methods study*. Saarbrüken, Germany: VDM Verlag Dr Müller.

Isaacs, T. (2008b). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English speaking graduate students. *Canadian Modern Language Review, 64*(4), 555–580.

Johnson, B., & Christenson, L. B. (2008). *Educational research: Quantitative, qualitative, and mixed approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Morley, J. (1991). The pronunciation component of teaching English to speakers of other languages. *TESOL Quarterly, 25*(3), 481–520.

Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign accented speech. *Language Learning, 48*(2), 159-182.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49*(s1), 285–310.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. London: Sage.

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

Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

Zielinski, B. W. (2008). The listener: No longer the silent partner in reduced intelligibility. *System, 36*(1), 69–84.

Suggested readings

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Harding, L. (2008). Accent and academic listening assessment: A study of test-taker perceptions. *Melbourne Papers in Language Testing, 13*(1), 1–33.

Johnson, B. R., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 64–91.

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

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

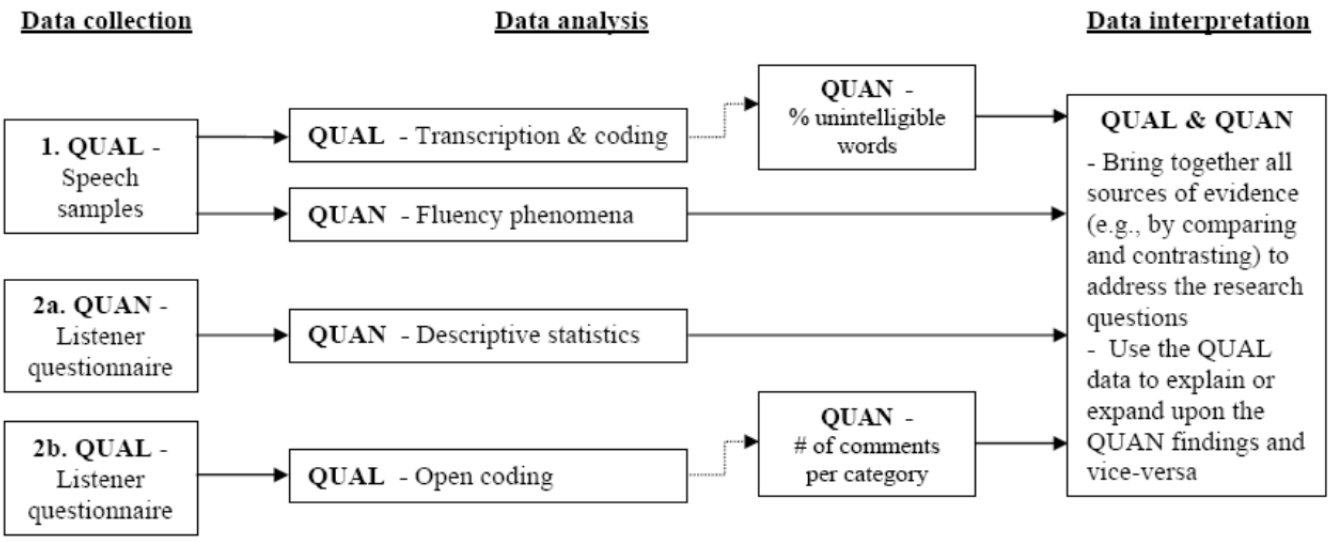


Figure 1. Mixed methods research design

Note. The speech data were collected prior to the listener data. Qualitative and quantitative data for the listener questionnaire were collected simultaneously.

Transcription symbols:
»speeds up«
_ brief pause
(1.) pause exceeds 1 second
‿linking
/ absence of linking
! abrupt detached articulation
≂ ≻ no inflection
• word emphasized in sentence
#consonant/vowel deletion
superscripted high pitch
superscripted indented H→L high rise intonation
blue color-coding intelligible word, even if pronunciation is unnormative
red color-coding unintelligible word, or speech feature contributes to unintelligibility

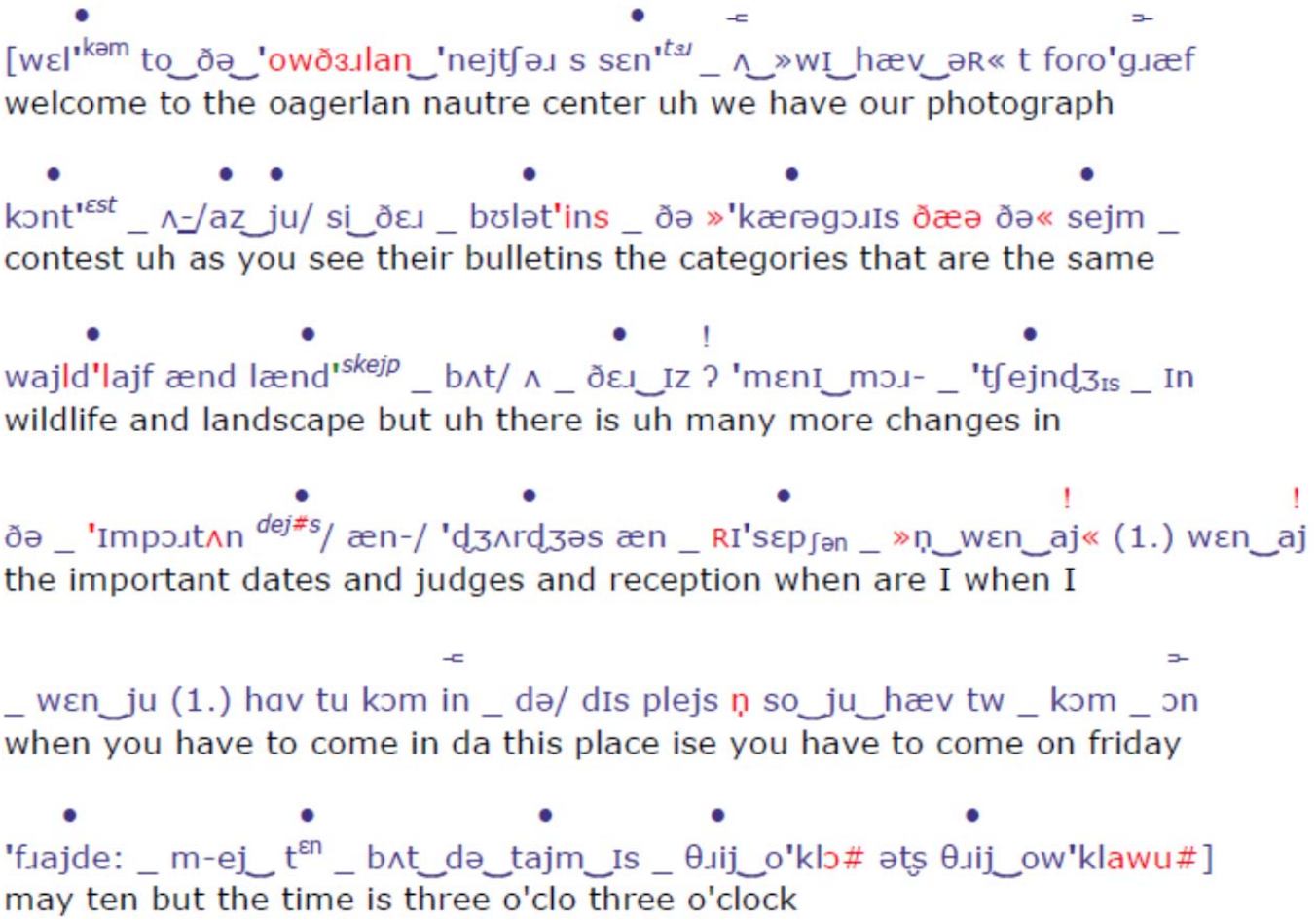


Figure 2. Sample excerpt of color-coded speech.

< FIGURE 2 DESIGNATED FOR ON-LINE VERSION OF ENCYCLOPEDIA ONLY>

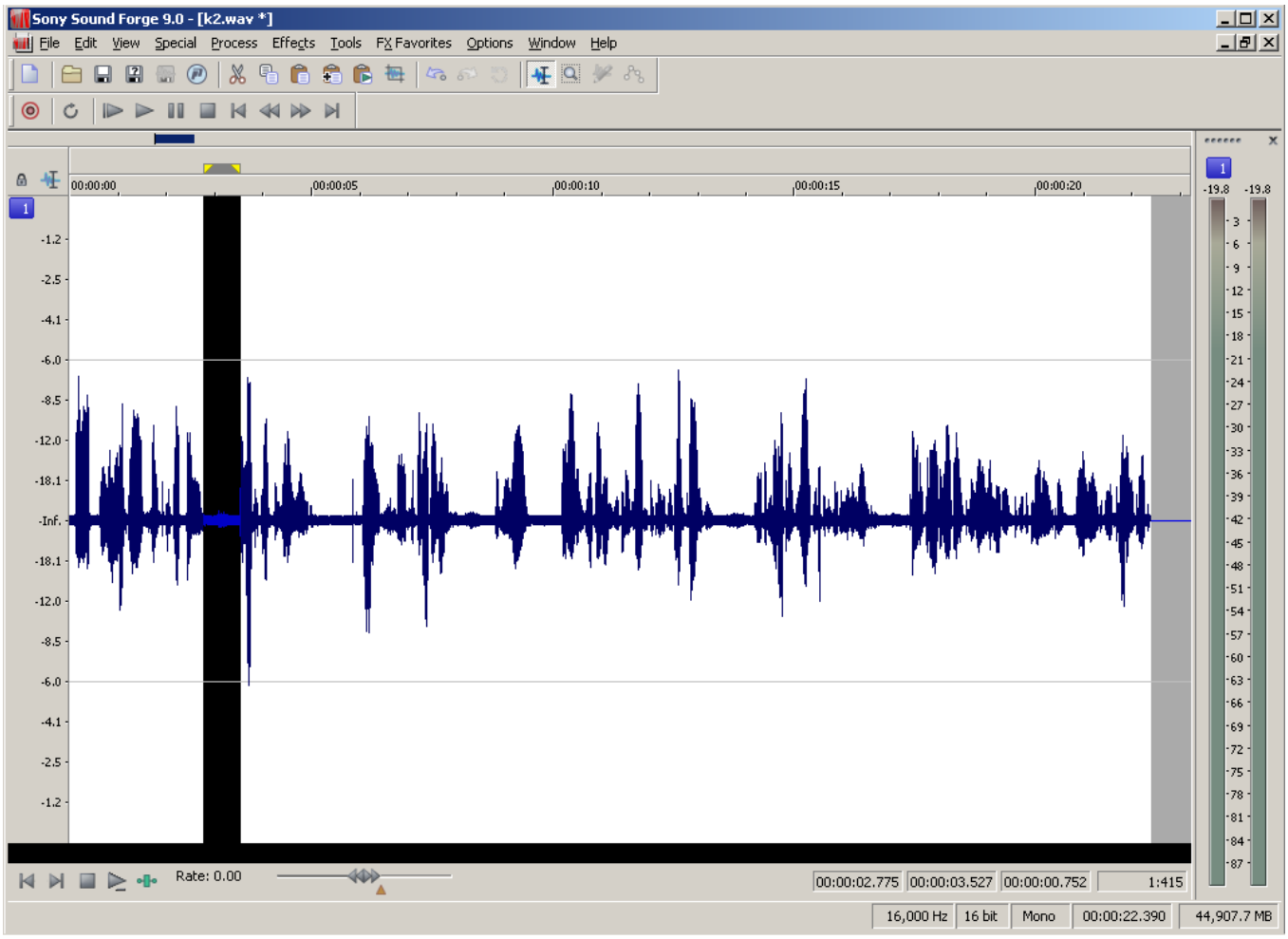


Figure 3. Screenshot of the quantitative speech analysis using *Soundforge 9.0.*

Note. The black highlighted area shows the measurement of a silent pause (i.e., as selected by the computer curser), which was used to derive the total pause length in the calculation of articulation rate.

<FIGURE 3 DESIGNATED FOR ON-LINE VERSION OF ENCYCLOPEDIA ONLY>

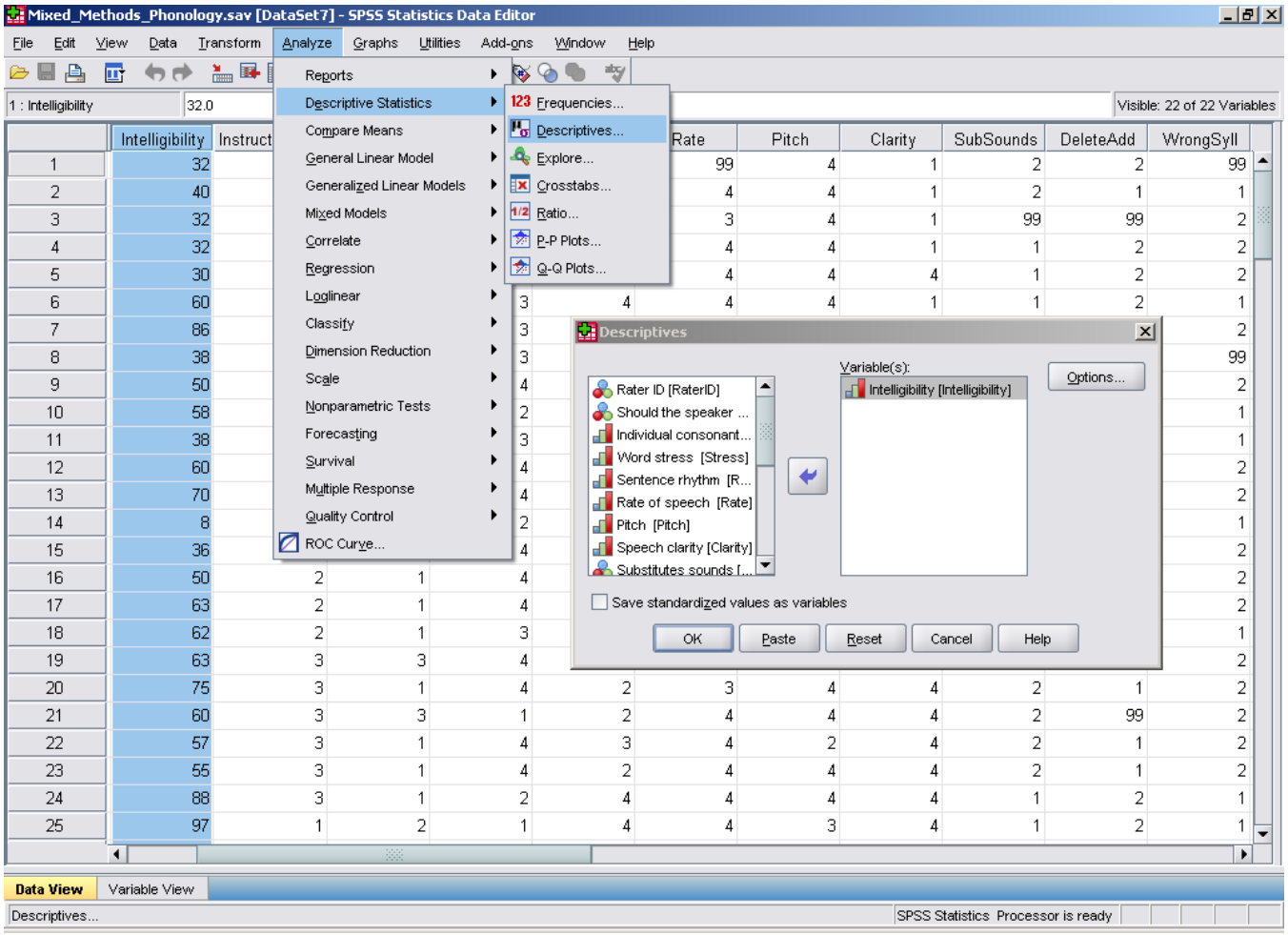


Figure 4. Screenshot of quantitative analysis of listener questionnaire using the software, *SPSS 17.0.*

Note. The blue highlighted column, "intelligibility," is being analyzed using descriptive statistics. In order to calculate intelligibility statistics for each individual speaker, it is necessary to carry out the following initial step. First, select the "data" option on the menu, then "split file." Next, change the option to "compare groups" and input the variable "speaker ID" in the textbox. Having done this, requesting descriptive statistics for intelligibility or any other statistics will yield output for each speaker.

<FIGURE 4 DESIGNATED FOR ON-LINE VERSION OF ENCYCLOPEDIA ONLY>

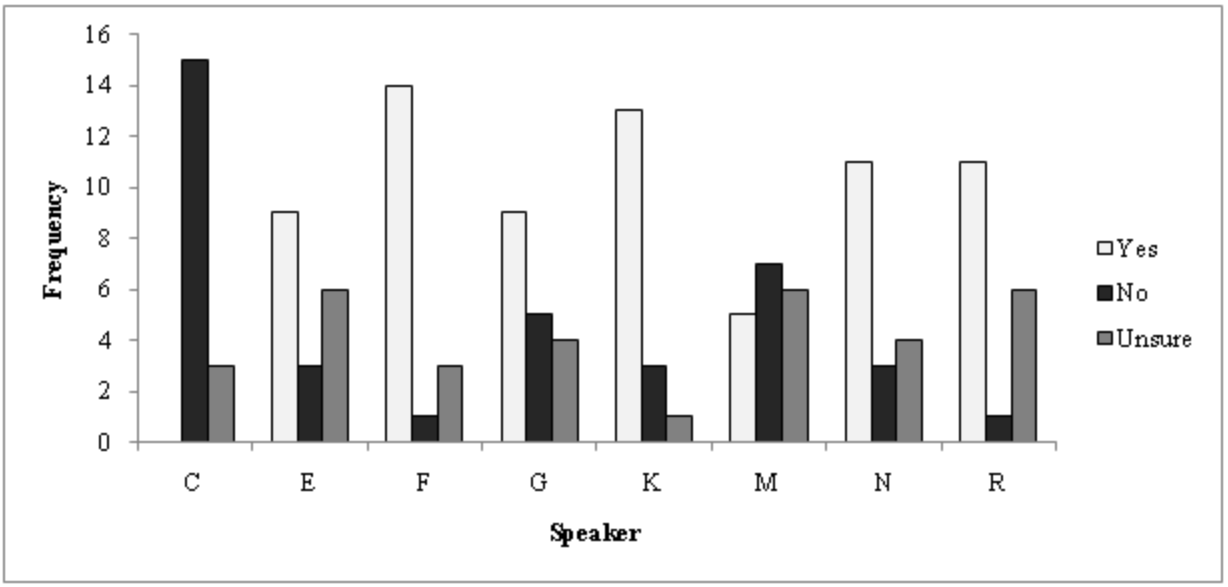


Figure 5. Frequency of listeners' responses as to the "TA question."

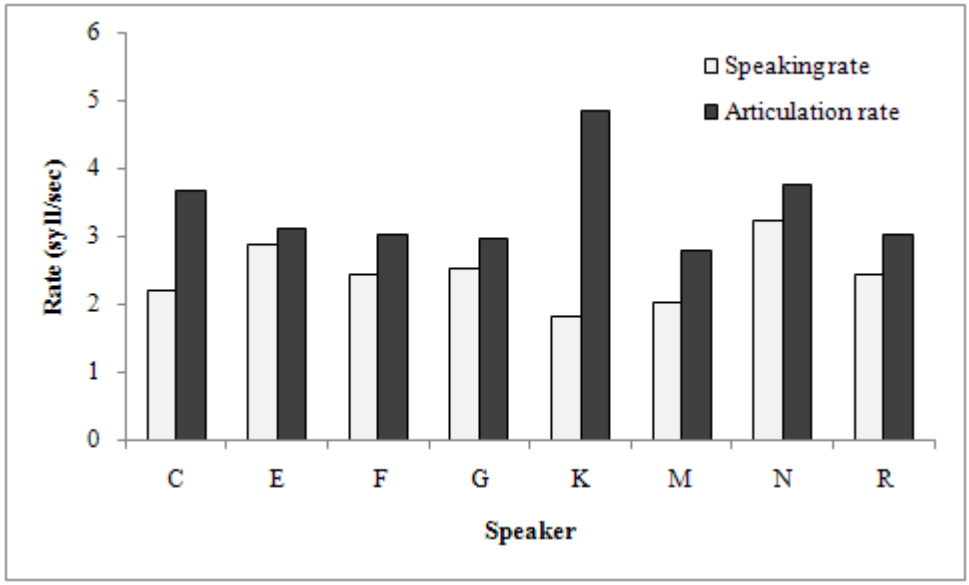
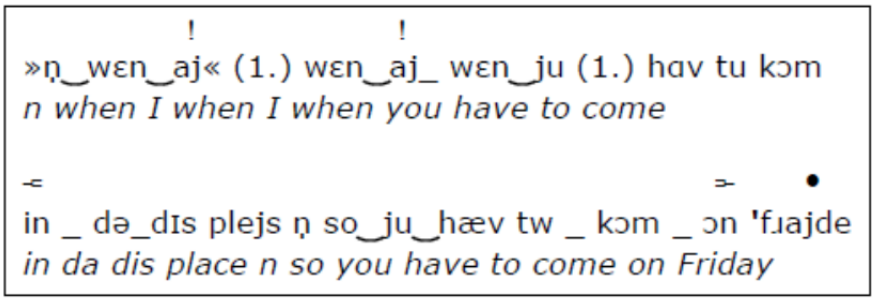


Figure 6. Speaking rates and articulation rates.

Isaacs, T. (2008). *Towards defining a valid assessment criterion of pronunciation proficiency in non-native English speaking graduate students, 64*(4), Figure 3, p. 570. Adapted by permission of University of Toronto Press Incorporated (www.utpjournals.com)



Transcription symbols:
»speeds up«
‿linking
! abrupt detached articulation
(1.) pause exceeds 1 second
_ brief pause
-= =- no inflection
• word emphasized in sentence

Figure 7. Example of an instance of unintelligibility in Speaker C's speech.