CHAPTER III

Methodology

Preparatory to generating a method for assessing the accuracy outcomes of a sample of complex codas, it is useful to review the objectives and what has been already been concluded.

- An examination of difficulty in complex coda production necessarily involves a description of repair strategy or error. (Sections 2.3 et seq.)
- An estimation of the degree to which errors may affect intelligibility is synonymous with an assessment of error gravity. (Section 1.4, 2.3.1)
- An assessment of coda accuracy should derive from a combination of the number of errors associated with the production of the coda and the gravity of such errors. (Section 2.3.1)
- A ranking of complex codas by their degree of mean accuracy would indicate the degree to which the codas could be improved.
- A hierarchy of instructional benefit would follow from a combination of coda frequency and coda accuracy, e.g., those codas used most often which could stand the most improvement should be taught first.

3.1 Intelligibility and Coda Well-Formedness

This study is specifically concerned with an evaluation of the complex coda and not with the pronunciation of word-tokens in their entirety—with errors in the

onset or vowel, or in syllable stress—any one of which might easily lead to a word seeming unintelligible to a given listener. It is assumed that the pronunciation of all complex codas which are alien to the Thai phonological system is a learned skill, which can be transferred to other words of the same coda type, provided there are not additional impediments, such as vowel diphthongs (see section 2.1.2). Thus, if a learner can pronounce the /ks/ coda in *box*, they should be able to pronounce the codas in the words, *fix* and *sticks* in equal measure, though they may well stumble over the /st/ onset in the latter word, which might indeed impact the coda in a mistake. But the pronunciation of the /ks/ coda will be generally acquired. Thai Learners will probably not be able to correctly pronounce the word *likes* because of the antecedent vowel diphthong—until they have mastered such combinations.

As unintelligibility is caused by a word or sequence of words which are outside a range of familiarity for a given auditor, it cannot easily be gauged by an evaluation exclusively of the coda of single word tokens. Moreover, there are common NS deletions and assimilations in connected speech (Section 1.4.2), which would tend to further complicate any attempt to link or correlate the well-formedness of (narrative) codas with what could be considered intelligible speech. Nevertheless, the same NS who practice connected speech elision can and do pronounce the complex codas arising from the grammatical **–ed** and **–s** inflections in authentic speech. There is also reason to think there may be a strong correlation between the well-formedness of codas and the overall intelligibility of the learners who utter them, as skills are usually learned in tandem with other skills—as evidenced by numerous scales and rubrics for which given levels of grammatical and lexical development coexist with similar developmental levels in pronunciation.

3.1.1 Rationale for Error Type as Opposed to Pass/Fail Assessment

How might individual codas be assessed in a manner which would likely correlate with the overall intelligibility of an utterance, and which would reflect the actual pronunciation errors a learner makes? Two basic options are available for scoring: pass/fail (0 and 1 point scoring), or a range of possible scores from 0 to 1. The difficulty with binary pass/fail scoring lies in determining whether the assumed token word is related to the gloss, and if it is, whether it might be intelligible. Scoring by error type avoids this difficulty, as words or tokens may be given partial credit if they agree with the gloss stem or in other ways resemble the gloss word sufficiently not to cause a communication breakdown. Evaluating error types also has the benefit that they may be tabulated to show systematic tendencies on the part of speakers, whereas failures would be assumed to be equivalent to one another. As much as arriving at numerical token intelligibility scores for a particular group of listeners might be relevant, it is more useful to know the specific tendencies of learners and to be aware of their strategies in dealing with awkward complex codas.

This would necessarily involve an analysis of speech, rather than a tabulation of listener responses—whether pass/fail or multiple choice. For example, the latter might involve possible answers to a given word pronunciation for the word, *judged*. Possible choices might include "just", "jut", "Jud", "judge", and the correct gloss word. The problem here is that not all possibilities are accounted for in these five choices. The word may instead be altered to "juts", or even to [tʃɛtʃəʔ] ~"chetch", as was actually the case in one sampled utterance. The actual utterance may fall through the cracks between the available word options an evaluator would devise to appear on a multiple choice test to be filled in by auditor-evaluators. Thus, a multiple choice test would less accurately describe what was actually said and which errors were made. It

might relate more to what was heard by an auditor, since most listeners seldom think in phonetic terms, but there would be the danger that the multiple word choices would suggest an answer that the auditor would not come up with on his own. *Judged* was guessed as 'Chesire" and as 'gesture' by the two auditors when pronounced as [tʃɛtʃəʔ]. It is very unlikely that a multiple choice test would include these two possibilities, partly because of the altered vowel. In any event, it seems more logical in such an evaluation to allow the auditors to arrive at their own guessed words, even if this might take more time.

Having decided to evaluate intelligible deviations (and unintelligible failures) by error-modification type, there are two possible ways to proceed to an evaluation of how well or how poorly the various coda types are pronounced: classification of the error(s) and the assignment of a numerical score for each pronunciation, based upon acceptable criteria for such scoring. The virtue of the former method lies in the absence of scorer bias in weighting the deductions for each error type, which might seem arbitrary to a disinterested observer. That is, the neutral observer might assign different weightings (deductions) for each error type than the evaluator, or s/he might assume that the same error type, e.g., deletion, might have different intelligibility outcomes, based on word environment, such as noted in section 1.4.2. The drawback of relying on error classifications alone is that complex codas may be subject to more than one error type, indeed, often suffer two or more concurrent errors which may be different. For instance, a complex coda token might experience one deletion of a consonant and one substituted consonant. To which error classification should the token be assigned: *Deletions*, or *Substitutions*? Clearly it should be assigned to both, but then a two-dimensional scale would result. The virtue of assigning a numerical score is that codas may be ranked by difficulty and easily compared with other codas

using a one-dimensional scale. If it can be shown that different score outcomes correlate with intelligibility at the narrative or discourse level, then the scores might be relied upon to the extent of this correlation. There are no failsafe evaluation methods for pronunciation. Scores are not perfectly valid measures, and should be taken as general guidelines, provided that the error classifications are also used. Indeed, it is the latter which are most useful as signposts for what is to be corrected. So while this study arrives at individual coda scores, it ultimately evaluates codas by error type, and tries to use the observed errors in devising proposed remedies.

A fundamental difficulty with an assignment of scores, or even a classification by error type (deletion, epenthesis, substitution, etc.) is that, despite the latter's prevalence in the literature (Section 2.3.1), the focus is on the speaker, as opposed to the listener. The attempt to be objective carries its own dangers, as the listener cannot be entirely factored out, as has been previously noted in section 1.4. While the listener cannot be factored out in spontaneous real-time speech, trained auditors can approximately determine what has been said through multiple re-hearings of recorded speech when provided with appropriate software programs that show acoustic waveforms. Since one objective of the study (Section 1.6) is to rank complex codas by their difficulty, much effort is expended on discovering what the informants actually say, and this perforce requires a classification by error or modification type. There may be myriad minute deviations from what the auditor might consider to be 'correct' or 'normal' speech, yet the practical results of an evaluation must fall within established parameters. Was a consonant omitted, was it substituted with another, and if so, was the substitute phoneme related? Were extraneous vowels inserted, or was the order of consonants reversed or changed? All of these modifications are attested to in the literature, and their development in learners and effects on intelligibility are

known to an extent (Sections 2.3, 2.3.1). The present study does not assume that there will be one dominant error type for all codas and informants. There may be different errors that are apt to occur depending on the phonological requirements of a particular coda for a particular individual and in a particular context, and overall dominant strategies or errors for each coda type. This information may be useful to the pedagogue or teacher in prioritizing the learning of the **–ed** and **–s** inflections and the codas they contain.

While a belabored rehearing of words or phrases, which an identification of phonemes and learner errors from recorded speech necessarily entails, is not synonymous with authentic listening, neither is the material used in the test—wordlists and narratives--'authentic'. If anything, it might test reading skills more than speaking skills among beginning-intermediate learners, and this is why more advanced informants who are confident readers and are presumably familiar with the words were selected for this study. Nevertheless, the test is also administered to a less advanced group of informants to see if similar patterns in coda accuracy are evident. The instrument must be artificial (or synthetic) in order to capture a wide enough spectrum of complex codas. The evaluation does, however, include subjective and personal assessments of the overall oral proficiency of the informants, as well as of words in context and in isolation.

3.1.2 Intelligibility of Codas Correlated to Comprehensibility of Speech

To validate the usefulness of individual coda scores, they are correlated with the overall intelligibility of the informants who utter the codas. While the individual complex codas can be transcribed and scored, the overall comprehensibility of the speech of the informants in this study, as reflected in the narratives and interviews,

may only be roughly graded. (Celce-Murcia, et.al., 1996) list several pronunciation rubrics, such as the ACTFL Proficiency Guidelines for Speaking, on an 11-point scale, with 3 novice, 3 intermediate, 2 advanced and 2 superior grades; however, the criteria are mainly functional and lexical, rather than geared to pronunciation per se. As the informants for the study read a script which contains the target complex codas, rather than create their own language, the six-point scaled Speech Intelligibility/ Communicability Index (Morley, J. 1994), reprinted in Celce-Murcia, et.al., (1996) is used (Appendix F).

Details of the evaluation methodology used appear in the following sections:

3.2 Informants and Auditors

3.3 Elicitation Instrument

3.4 Data Collection Procedure

3.5 Data Analysis Procedure

3.2 Informants and Auditors

It is important that the informants participating in the evaluation manifest typical Thai traits (see section 1.1) in their transference of Thai phonology to English yet be able to articulate to some extent the various complex codas by their familiarity with the English words used in the test. To meet the first requirement, they should not have had *recent* or *extensive* naturalistic exposure by residing in English-speaking countries which might have altered their interference patterns. To meet the second, undergraduate students who are enrolled in the same program within the International College at Payap University are evaluated. Thus, the informants were expected to be fluent English speakers who still have a Thai accent. To be representative of the Thai

educational system in Chiang Mai, they should have completed at least six years of secondary English instruction in Thai, bilingual, or international schools. Four informants were chosen to provide a wide enough sample, yet keep the token count manageable.

Auditor-evaluators who were educated NS TESOL teachers were chosen to represent a segment (North American) of NS likely to interact with local Thais in English. However, the auditors probably have more exposure to local Thai accents than does the average American (or European or Chinese) tourist or business customer who visits Chiang Mai. This added exposure coupled with their language training would indicate that the auditor-assistants should recognize a wider than average spectrum of speech as intelligible, and also be able to more accurately transcribe it.

Table 1 below presents educational profiles and other individual characteristics of the four informants who participated in the study. The personal data was drawn from the interviews the evaluator held with each informant prior to the complex coda assessments. The interview instrument is described in more detail in 3.3.4.

TABLE 1: INFORMANT HISTORY

Criterion	Informant 1	Informant 2	Informant 3	Informant 4
Schools attended when learning English	English instruction in Thai schools P1-M6; AUA and British Council for two years; Payap International College for 3 years	Thai school for M1 to M3; Prem International School for M3-M6; 1 year at Payap International College	Singapore M1-M2; American School of Bangkok (ASB) M3-4; Grace International, M5-6; Payap International College (PIC) for 3 years	Many different High Schools in New Zealand; Prem International M5-6; Payap International College for 1 year

/Continued

Table 1—Continued				
Criterion	Informant 1	Informant 2	Informant 3	Informant 4
Years abroad, when and where	1 year exchange student in 11th grade in U.S.A. four years ago	2 months in the U.S.A. as a tourist four years ago	Singapore for two years as a student; returned 7 years ago	New Zealand for 3 years as a student; returned two years ago
Opportunity to practice English with native speakers	Occasional opportunity, as she has some native speaker friends	Some opportunity to practice English in school (Prem and Payap), few native speaker friends	Was one of the only Thai students at Grace; many native speaker friends. Many friends at PIC	No native-speaker friends, but father wants her to learn English, as it's the "world language"
Other remarks	Since returning to Thailand four years ago, she feels her "English seems to be getting worse".	Classes at Prem International had about 20 students of widely differing nationalities. No single accent prevailed	Little opportunity to speak English at ASB; Grace International was "a direct line" for her to speak English	Immersion on English classes in New Zealand was difficult; attended a language school for 3 months in preparation

Two native-speaking assistants, who are both TESOL teachers enrolled in the graduate TESOL program at Payap University, volunteered to audit the recorded narratives and wordlists of the informants and to assist with their evaluation. Auditor 1 was born in the U.S. of native-speaking parents, and grew up in an affluent and predominantly non-immigrant suburb of San Francisco, California. He has taught English for one year in Mexico, and is fluent in Spanish. He has resided in Chiang Mai for 1 year, and is familiar with the Thai accent. Auditor 2 was born in Guyana, South America of Indian bilingual English speaking parents and has attended school

for many years in the eastern U.S., finally attaining advanced degrees in Psychology. He has also resided in Chiang Mai for about 1 year, interacts daily with Thais, and coincidentally also is fluent in Spanish. Thus, the linguistic backgrounds of the two auditors are reasonably similar: both are educated and have been exposed to more than one other language besides English. Both are more familiar with North American than with British or other regional NS dialects. Auditor 2 is about 15 years older than auditor 1.

3.3 Elicitation Instruments

The elicitation instruments employed the following:

1. Wordlist test-words (tokens) in isolation
2. Narrative test-words (tokens) in isolation
3. Narrative for overall impressions of intelligibility
4. Interview for overall impressions of intelligibility

3.3.1 Derivation of the Test Sample of Words

There are about 150 existing ending consonant clusters in English where /r/ is a constituent (see Appendix C), and about 115 where it is not (see Appendix B). In order to bring the test load down to a manageable number which would be representative of normal spoken English and still readily evaluated, various criteria were used. These were: common English native speaker reductions, rarity, difficulty (phonetic markedness), and a consideration of the Thai language's phonological constraints.

Effective realizations of complex codas can be affected by phonological processes similar to those in connected speech (Celce-Murcia, et. al. 1996), namely,

reduction or simplification (deletion of a consonant) and addition (consonant intrusion). The former process occurs in the initial listing of consonant clusters in the following forms, which are all triplets or quadruplets: /ndz/ ("hands"), /kts/ ("products"), /nθs/ ("months"), /fts/ ("gifts"), /mpts/ ("attempts"), /ŋkθs/ ("lengths"), and /ksts/ ("texts"). In all of these cases the medial alveolar or inter-dental stop or fricative (/t/, /d/, /θ/) is likely to be omitted even in careful speech by native speakers. There are other examples of the reductive process, which could be listed, such as /pθs/ ("depths") and /sts/ ("costs"). The difficulty for Thai and other speakers of the dental fricative, /θ/ would tend to disqualify the former cluster from consideration, while the latter cluster presents difficulty because /s/ occurs twice and is homorganic with the medial consonant, /t/.

Addition (consonant intrusion) occurs after "ng" when the /k/ is naturally inserted after the /ŋ/ and the following dental fricative, /θ/ ("length", lɛŋkθ), and when a plosive is inserted before a voiceless dental fricative (/θ/) ("warmth", /wɔ:mpθ/; "month", /mʌntθ/). An inconsistency arises with the plural forms ("lengths", "months"), where the medial dental fricative is again likely to be omitted due to the pronunciation difficulty; we get / lɛŋks/ and /mʌnts/. Here we have the simultaneous operation of deletion and intrusion. For purposes of this study, "students" (/stju:dənts/), "sense" (/sens/), and "months" (/mʌnθs/) are all assumed to have an equivalent ending consonant cluster (/ns/), due to the reductive and additive processes mentioned. Other equivalent sets include "hands" (/hændz/) and "cans" (/kænz/), "products" (/prɒdʌkts/) and "ducks" (/dʌks/), "scripts" (/scrɪpts/) and "ships" (/ʃɪps/), "lofts" (/lɒfts/) and "scoffs" (/skɒfs/), "attempts" (/ətɛmpts/) and

"camps" (/kæmps/). "Lengths" and "links" would differ ever so slightly only by virtue of their quite similar vowels. Linguistic purists would perhaps object to the homophonic process of the foregoing, yet even careful speech of native speakers involves these elisions (Jenkins 2000:142-3; Celce-Murcia, et.al.1996: 83, 88). Surely it would be unreasonable to expect Thai learners uneasy with English complex codas to improve on standard NS practice?

The entire set of complex codas (initial listing of RP codas) has been further reduced for two more reasons. Although some complex clusters may be relatively easily pronounced, e.g., "glimpsed" (/glɪmpst/), they occur so infrequently that their accurate production is rather unimportant. "Clothes" (/kləʊðz/) contains a double ending consonant cluster, while "glimpsed" contains a quadruple cluster, yet the extreme difficulty in the former word of pronouncing a voiced inter-dental fricative in final position with another voiced consonant following (even for native speakers) would seem grounds for its removal from consideration in the list of viable word test-candidates.

Some clusters are so frequently used that they are tested, regardless of the pronunciation difficulty (/lz/, θs/, /sks/, /skt/, /lvz/, /lvd/, /lts/, /lpt/, and /kst/). The least frequent double cluster in the final list (Appendix D) is "belonged" (/ŋd/ cluster, 0.06/1,000 words); the least frequent triple cluster is "films" (/lmz/ cluster, 0.05/1,000 words). No quadruple ending clusters are tested, as they are either frequently simplified (reduced) by native speakers, e.g., "attempts" /mpts/, "instincts" /ŋkts/, or are very rarely used, e.g., "glimpsed" /mpst/, "waltzed" /ltst/, "jinxed" /ŋkst/.

Consideration of whether to use a North American rhotic view of consonant clusters, where /r/ would be a valid constituent, or to use a British RP (Received

Pronunciation) view, where /r/ is omitted from the ending consonant cluster was finally resolved by Thai phonology. Thai is a non-rhotic language, and the Thai /r/ can only occur pre-vocalically. Even here, it is often pronounced as a trill instead of as an approximant. Hence, Thai accords better with the British standard than with the North American, especially as the Thai vowel set can approximate to the RP vowel set (including diphthongs). British RP is the preferred model, and its phonemic representations are used in this study.

Syllabic allomorphs belonging to both **–ed** and **–s** inflections ostensibly could also be represented, provided that they also contain complex codas; it makes little sense to compare the pronunciation of syllabic allomorphs not containing consonant clusters with non-syllabic allomorphs that do if it is presumed that consonant clusters are the prime impediment in speech. Unfortunately, there are no practical examples where syllabic allomorphs could be added to consonant clusters and would need to be pronounced as consonant clusters. For example, /nt/ and a sample word, *dent* inflects with the syllabic **–ed** allomorph to *dented* /dentəd/. The problem here is that the consonant cluster /nt/ can then logically be segmented into the two syllables /den/ and /təd/, rather than /dent/ and /əd/. With triple consonant clusters that still can be inflected, e.g., *attempt* /ətempt/, the inflection yields *attempted* /ətemptəd/, and the medial /p/ is apt to be elided. We get instead /ətemtəd/, which can be segmented into three syllables: /ə/, /tem/, and /təd/. Hence, there really are no suitable candidates to test in their syllabic forms which would remain ending consonant clusters.

3.3.2 **Wordlist** (Instrument 1)

Fifty different complex codas are represented by 50 different words (Appendix D). These are the most common complex codas resulting from the grammatical **–ed** and **–s** inflections, and they individually occur at least 0.05 times per 1,000 words. There are 6 voiced /d/ and /z/ simple codas following the diphthongs, /aɪ/, /aʊ/, and /ɔɪ/, which are usually analyzed in Thai as already containing consonants, viz., /aj/, /aw/, and /ɔj/; 29 doubletons containing 9 voiced /d/, 8 voiced /z/, 6 voiceless /t/, and 6 voiceless /s/ codas; and 15 tripletons containing 2 voiced /d/, 2 voiced /z/, 6 voiceless /t/, and 5 voiceless /s/ codas. Frequencies in English usage vary from the most common /nd/ coda, which occurs 31.7 times in 1,000 words; to the /lmz/ coda, which occurs 50 times in 1 million words. The entire set of 50 codas occurs 83.35 times in 1,000 words, or slightly more than once in twelve words in overall English usage.

Polysyllabic words were used (twice) when they were the most common representatives of their coda type and were deemed more familiar to the informants. As research in *Natural Phonology* cited in section 2.1.5 has indicated, monosyllabic words are more perceptually salient to young L1 speakers than are bi-syllabic words; this may be analogously true for adult L2 English learners.

3.3.3 **Narrative** (Instruments 2 and 3)

Studies cited in section 2.3.1 have shown that the most formal oral tasks, such as reading wordlists, manifest different degrees and types of pronunciation errors than do less formal tasks, such as reading a narrative from a text. Thus, two different tasks were arranged to capture a wider spectrum of speech functions, and hence, a more

realistic picture of coda accuracy outcomes. The narrative (Appendix E) is viewed in two different ways: segmentally for coda well-formedness (instrument 2), and at the phrase-clause level and in its entirety for overall intelligibility (instrument 3). The narrative is prefaced by an introductory sentence which contains few complex codas, allowing the informants to ease into the reading of the narrative. There are two versions of the *Bird Flu* narrative which present the sentences in different orders, and these were evenly distributed amongst the informants. Two doubletons (*besides* and *described*) follow a diphthong, and a research question is to determine whether the prevailing repair strategy is to simplify the complex coda, alter the vowel it follows, or both. There are 12 plural nouns (*cats, beliefs, boys, cows, chickens, things, jobs, eggs, birds, risks, deaths,* and *films*) representing /s/ and /z/ codas, 6 third-person singular present tense verbs (*thinks, makes, tells, seems, gives,* and *solves*), and 3 words that could fit both or neither category (*steps, results, besides*). Another research question: does morpheme status affect accuracy?

While it incorporates the same 50 different complex codas as the wordlist, in the attempt to measure their accuracy of production, the narrative also attempts to mix formal ("The Thai government has managed to adopt a change …") and informal styles ("The boys on the farm were used…") of speech. On balance, the style probably leans toward a formal style of speaking, which may be more appropriate for adult university students than for other learners. Certain tokens probably demanded a more formal style: *managed, adopt, solves, solved, described, launched.* The narrative does not purposely avoid potential connected speech contexts, which are noted in the analysis. Many of the **–ed** inflections are embedded in the present perfect tense, and so grammatical context is provided to the informants who read the passages, as well as to the auditors who hear them.

A large number and wide variety of complex codas are tested, so evaluating authentic oral communication was not feasible, except for general impressions of intelligibility in the interviews. In any event, learners often employ coping mechanisms to avoid using complex codas in authentic communication by substituting other words which do not contain complex codas or by redundancy in order to get their meaning across (Section 1.2). It would be nearly impossible to 'force' the use of all, or even most, sample complex codas even using prompts. Celce-Murcia, et.al., for example, contains two prompted exercises—one for **-s** third person singulars, and one for **–ed** past tense in which the students choose the correct regular verb forms and pronounce them. These exercises would allow for the use of only 8 **–s** non-syllabic codas and 9 **–ed** non-syllabic codas. At a certain point the attempt to be authentic becomes too obviously transparent to the informant, and the exercise starts to lack authenticity. It becomes entirely predictable what the evaluator is looking for. Such a prompt actually tests recall of the inflected forms. Reading paragraphs aloud involves a different cognitive load than asking the informant to read individual words or to recall inflected words, partly because of the time element, and partly because of the demand on grammatical knowledge. All three are distinct activities which differ from authentic speech. So it should be understood that the accuracy of the codas as measured in this elicitation instrument is only partly related to authentic speech pronunciation accuracy. Perhaps it is equivalent to how accurate the speaker might be if asked to repeat a word which was not understood.

3.3.4 **Interview** (Instrument 4)

Authentic speech was evaluated only once, in the interviews, and this was supplemental to the evaluations of the whole narratives. Interview content was not

uniform to all the informants, but varied depending on what questions were asked by the interviewer, and how comfortable the informant was answering the questions. While the general nature of the interview questions remained the same for all four informants, follow-up questions differed depending on the answers given by the informants to the original questions. Thus, some informants were relatively talkative, and some more reserved.

The formal elicitation instrument used to generate each informant's coda production (described in 3.3.2 and 3.3.3) was prefaced by an interview, lasting an average of three minutes, where the student was asked a variety of questions relating to background: schools attended, academic and career interests, opportunity to practice English and in which settings. This information is contained in Table 1. In addition to the discovery of informant history relating to education and the use of English, the interview was employed to get acquainted personally with the informants and to generally loosen tongues and put informants at their ease for the subsequent wordlist and narrative.

3.4 Data Collection Procedure

The informants were not told the exact reason for the recording and evaluation of their speech, only that they were encouraged to express themselves naturally and as accurately as possible in the narration, as if they were broadcasters giving a news report or business-persons an opinion of current events. Any sentences containing mistakes that the informants were aware of making were repeated once if desired. The intention of the test was to encourage natural, but careful speech, such as might be used when giving a presentation. The wordlist was always given before the narrative as preparation. The words from the wordlist were presented separately on individual

flash cards. The wordlist took about 5 minutes per student. A short break followed the wordlist. The wordlist was pronounced and recorded in reverse order by the second student of each pair to lessen the tendency of the data to cluster uniformly as it was audited in the same order it was recorded. Thus, range and frequency effects, as described in Parducci (1965) were minimized.

3.5 Data Analysis Procedure

The wordlist and narrative recordings were segmented into lexical and phonetic chunks with the aid of *Speech Analyzer* and *Phonology Assistant* (Speech Tools 2.2, SIL International, 2005). After the narrative and wordlist test words in isolation had been heard and (orthographically) transcribed by the auditors without their having seen the gloss text, one unedited narrative was played back in phrasal or clausal chunks (about 6 seconds in length) 3 to 5 times for one auditor to transcribe. Individual complex coda tokens were evaluated in the order in which they appeared in each narrative. The auditors took breaks as desired. On a following day the same process was repeated. There were thus about four initial sessions per auditor, and the gaps between sessions were designed to limit the memory of the auditor as to what the narrative or wordlist might contain. The evaluator then returned to the phonetic segments and made changes as indicated to reflect the perceptions of his two auditor-assistants on the isolated words. Naturally, perceptions of the auditors which were affected by grammatical considerations in the unedited narratives were ignored for the phonetic transcriptions. Auditors were instructed to use grammatical/contextual clues solely in the un-segmented narratives (instrument 3). Then the cycle was repeated, with the auditor-assistants this time listening to the individual segments and helping the evaluator decide whether a coda consonant was articulated, or whether it was

substituted with another consonant or deleted. Whole narratives and interviews were evaluated last to arrive at scores on the 6-point intelligibility scale (Appendix F). They were not used to generate scores for individual complex codas. Only instruments 1 and 2 were used for this purpose. Formal instruction to the auditors are in Appendix J.

Correlation between the individual scores of the token codas and the overall intelligibility of the informants who uttered them employed instruments 1 and 2 and instruments 3 and 4, respectively. Instruments 1 and 2 were used simply to determine the basic intelligibility of the informants as they relate to the well-formedness of the individual coda types. The focus of the evaluation started from the least comprehensible instrument 2 of the least comprehensible speaker (in the opinion of the evaluator) and worked its way towards the most comprehensible (instruments 3 and 4 of the clearest speaker). It also alternated between the wordlist and narrative, discouraging auditor familiarization with the test tokens. Instruments 1 to 3 were performed first for all the four informants, and instrument 4 was saved for last and performed on all 4 informants at the same time, with numerous re-hearings for comparison.

3.5.1 Example of an Analyzed Narrative and Audition Styles

The narrative of informant 4 (Section 3.2) was auditioned first by the two auditors separately (on different occasions) for a comparison of their listening skills. Auditor 1 evidenced fewer and shorter digressions from the intelligible message of the narrative, and his top-down approach required fewer and longer phrasal repetitions than did Auditor 2's more bottom-up approach. Nevertheless, some similarities are apparent. The **–ed** and **–s** inflections are lacking: 14 **–ed** and 29 **–s** complex codas are

missing in the former transcription, while the second transcription shows 18 **–ed** and 23 **–s** complex codas are missing.

Narrative of Informant 4

As heard by Auditor 1
Average phrase listening repetitions: about 2.5
Average phrase length: about 7 seconds

Key:

Incorrect word

[correct text]

I am a businessman in Thailand and I worry about bird flu. I worry that we are not doing enough about it. We watch[ed] film[s] that show[ed] chicken[s] burning [buried] alive in death camps. This make[s] me very afraid. Bird flu has changed some belief[s] about [amongst] many farmer[s]. The Thai government [has] manage[d] to adopt a change in farm policy and has moved to help some large farmer[s] to use bird flu test[s]. Why [While] the right question[s] will [were] ask[ed], few problem[s] will [were] solve[d]. Once we solve [results] our [of] research on bird death[s] show that cats, pigs, or cow[s] are linked to bird[s] we human[s] are at risk. But [the] government lie[s] about the problem and think this make[s] the cow [crowd] to content. The step[s] is [it's] taken haven't fixed much and farmers must be dragged into the modern world. They haven't even washed the egg[s] they have sold.

Because [The costs] of [the] bird flu can ['t] be described; beside[s] the risk[s] to insurance, tourists might avoid coming here. Job[s] that have helped to list [lift] many would be at risk. Thaksin tell[s] us things are fine, but at the end [the aid] he give[s] so [solves] nothing. The boy[s] on the farm will [were] use[d] a [nd] lie[d] too and it seem[s] they pick[ed] the wrong r [l]eader. True, they seem[ed] to have enjoyed the last five year[s], but many [are] camp[ed] out in Bangkok who have called for him to step down. Many who belong[ed] to the TRT last year has[ve] now launched[ed] attack[s] against him.

Narrative of Informant 4

As heard by Auditor 2
Average phrase listening repetitions: about 4.0
Average phrase length: about 5 seconds

I am a businessman in Thailand [and] I worry about Bird flu. I worry that we are not doing [enough about] it. We wash [watched films that showed] clean __ chicken[s buried alive in death camps]. This makes me very afraid. Bird flu has changed [some] beliefs about [amongst] many farmers. The Thai government [has] managed to adopt a serve [change] in farm policy it's impossible for [and has moved to help some] large farmers to use bird flu tests. Write [While] the right question[s] we'll [were] ask[ed] feel [few] problem[s] with health [were solved]. Want [Once} result[s] of research on birthday [bird deaths] sure [show] that cat[s], pick [pigs, or cows] align [are linked] to bird[s] we human[s] are agree [at risk]. But [the] government lie[s] about the problem and [thinks] this makes them [the] proud [crowd] content. Bus stop [But the steps] is [it's] taken haven't fix[ed] much. A flower [Farmers] must be dragged into the modern world. They haven't even washed the air [eggs] they have from [sold]. Because [The costs] of the [bird] fruit [flu] can['t] be described; beside[s] the risk to interest [insurance], tourists might avoid coming here. [Jobs that have] This help[ed] to this [lift] many would be at risk. Thaksin tell[s] us things are fine, but the aid he give so [solves] nothing. The boy[s] on the farm [were] use[d] a[nd] [lied] to and it seems he [they] picked the wrong reason [leader]. Too [True], they seem[ed] to have enjoyed the last five years, but when he [many are] came [camped] out in Bangkok they [who have] call[ed] for him to step down. Many who belong[ed] to the TRT have now launched attacks against him.

3.5.2 Derivation of the Scoring Metric

Section 3.1.1 presented a rationale for using a one-dimensional point scoring method, rather than a multi-dimensional error classification, in ranking coda production. The scoring metric in this study was devised after the data (the recorded speech) was gathered and some analysis of the data had taken place. It is thus an *ad*

hoc as much as an *a priori* scoring system. It attempts to weight error deductions based on both speaker-centered and listener-centered criteria.

Having decided to assign numerical scores to complex coda production, what general speaker-centered guidelines are available that might reflect the abilities and overall communicability of speakers who commit the various error types? As noted in Section 2.3.1, both Lin (2001) and Abrahamsson (2002) have found that more advanced speakers tend to choose epenthesis over deletion; moreover, learners at all developmental levels tend to increase their use of epenthesis and decrease their deletions in more formal tasks, even if this leads to a greater number of errors. More casual tasks tend to involve deletions, implying that it is epenthesis that is the *learned* strategy, and deletion, the *natural*, inherent strategy. Natural Phonology also posits that learners, whether adult NNS or NS children, tend to undergo an evolution from deletion, substitution, and epenthesis, to final correct form in their production of complex codas. Therefore, a speaker-centered evaluation methodology would broadly weight deletion errors more heavily than epenthesis or substitution errors, because the speakers who practice more deletion tend to be less experienced or focused beginners.

A listener-centered scoring methodology would make note of intelligibility or comprehensibility problems with specific listeners or an audience, and would make a limited number of revisions to the broad speaker-centered scores, based upon mismatches between the scores and the intelligibility results. Revisions would be 'the exception which proves the rule', rather than the new rule, itself, else there is a danger that the scoring scheme would be too complicated, as well as ad hoc. It would not be replicable. Justification for all revisions would have to be given, since the integrity, as well as the simplicity, of the scoring system is of primary importance. The scoring metric originated from a desire to establish a simple point scale which could handle

the various error types uniformly from a speaker-centered perspective, and was later refined ad hoc from the wordlist data, i.e., from the disparity between the gloss and the assumed word of the evaluators. Such a scoring metric clearly could benefit from continual refinements and adjustments, based upon correlations with intelligibility in word, narrative, and off-test (interview) speech, provided these refinements were logical and explained in detail.

There were relatively few words in the sample which were uniformly unintelligible for both auditors, particularly when there was context, as in the narratives[9]. This was due to many and variable factors. If the surrounding words presented little difficulty to the auditors or provided grammatical (e.g., present perfect tense) or collocational context (e.g., 'watched films'), then wrongly pronounced words were more likely to be guessed right. Errors or mistakes affecting the onset and vowel of a test token could cause unintelligibility even if the coda was pronounced intact. Unintelligibility showed up more in the wordlists where there was no context, or in the narratives when segmented into individual words.

What follows is a scoring method which attempts to evaluate errors from both the speaker's and the listener's point of view. Departures from the original speaker-centered point scale are noted as ad hoc adjustments.

3.5.2.1 Deletion (elision)

An error type such as deletion (elision) may often lead to unintelligibility, in the case of many first consonant deletions, may merely remove the grammatical **–ed** and –s inflection, e.g., of the final /t/ in *helped* [helpt →help] or the final /s/ in *beliefs* [bɪlifs → bilif], as in the case of most final consonant deletions, or may, as in the case of many medial or first consonant deletions, be a way of handling awkward codas

(/θs/, /sks/, /pts/ /sts/, etc.), which many native speakers also practice. So to seek a uniform point deduction for deletions in all three positions may seem illusive when they may have decidedly different intelligibility outcomes. The major categories of deletion, voicing substitution, and other substitution of phonetic feature (place or manner of articulation) were compared against each other, both for what they show about the limitations of the speaker (given the observations from *Natural Phonology* already cited) and their effects on intelligibility for the auditors, and most deletions were determined to be systematically more serious than feature substitutions, and the latter more serious than voicing substitutions. Thus, deletions provisionally rated a deduction of 0.4 points; feature/place of articulation substitutions, 0.3; and voicing substitutions, 0.2 points. The scores do not correlate perfectly with the intelligibility of the tokens, even when evaluated as individual words divorced from context. This is because tokens that had lost their inflections by deletion (and thus had a 0.4 deduction) were deemed potentially intelligible in context as long as the stem was maintained—as they often are in practice.

Deletion of the final consonant in the inflected words that constitute the test sample modifies past tense to present and plural case to singular, but the stem is unaffected, retaining some contextual intelligibility. Deletion of the first consonant in a complex coda (usually /l/) does affect the stem, and can seriously impact intelligibility. Medial deletions, mostly of /k/ in "risks", "linked", or "thinks", or of /p/ in "lamps" generally are not serious if they are the only deletions present, since in the first case the /sks/→/ss/ reduction is common with native speakers (Celcia-Murce, et.al.,1996), while in the other three cases, the medial consonant is an effective consonantal transition between the velar or bilabial nasal and the alveolar /s/ or /t/. Medial consonant deletions in "thinks" and "lamps" might be confused with "things"

and "lambs", but the latter are supposed to have voiced coda endings. It is only the chronic devoicing practiced by many Thai learners which might cause "things" [θɪŋz] to be confused with [θɪŋs] or "lambs" [læmz] with [læms]. The deletion of medial /dʒ/ in "changed" is more serious, implying "chained". Most deletions were given a deduction of 0.4 points. The least serious medial consonant deletions in /mps/, /mpt/, and /ŋkt/ have an ad hoc deduction of 0.2 points. Commonly used NS reductions of /θs/ and /sks/ →/ss/ and /skt/→/sst/ also have a deduction of 0.2, provided the /s/ is lengthened as compensation for the deleted penultimate consonant; otherwise the deduction is 0.3. The medial deletion in /ŋks/ has a deduction of 0.3 as the functional load of the medial consonant[10] is higher than with /ŋkt/, but the voicing status helps identify the coda as voiceless.

Some feature/place substitutions, e.g., /θ/→/t/ in "deaths" (with a 0.3 point deduction), may have been unintelligible (as the assumed word, "debts", was unrelated to the stem, i.e., "death"), yet a typical interlocutor could become aware of such commonly used systematic feature substitutions in due time, and adjust accordingly. This is even truer of voicing substitutions. But with deletions there is no opportunity for the listener to adjust. There is an irrecoverable loss of phonemic information. Moreover, a speaker who tends to delete in one context may delete in other contexts, where the result may indeed be unintelligible. This can be seen with uninflected forms of codas in the test, e.g., with the word 'act', as opposed to 'picked' (both involving the /kt/ coda), or the word 'crowd' as opposed to 'allowed' (both involving the diphthong coda /aʊd/). The deletion of the final coda consonant had

more serious repercussions in the uninflected tokens, where the result was often unintelligible: /æk/, /kraʊ/.

3.5.2.2 Substitution

Voicing substitution, e.g., of the final /z/ in "boys" /bɔɪz/ to /s/ [bɔɪs], is the most common form of substitution for those inter-languages lacking voiced finals. Depending on functional load, this error may affect intelligibility. With "boys" one alternative chosen by auditor 2 was "voice" /vɔɪs/, in this case retaining the voiceless final while substituting the bilabial plosive onset /b/ with a labiodental fricative [v]—two changes to the onset consonant, namely place and manner of articulation. With a word like "seems", pronounced [sims], the alternative chosen by the same auditor was "since" /sɪns/. This involved the substitution of the first bilabial nasal with an alveolar nasal and of the close /i/ vowel with a near close /ɪ/. In this example, the auditor assumed one substitution of two phonemes to guess "since", while the exact phonemic equivalent for /bɔɪs/, namely, the proper name, "Boyce", would probably not occur to the listener. This auditor was evidently sensitive to the voice status of the final consonant. But it is safe to assume that voicing substitutions, being the most common substitutions by far, present fewer obstacles to intelligibility than does substitution with an unrelated consonant, i.e., one that differs in place and manner of articulation. Voicing substitutions were given a deduction of 0.2 points. No ad-hoc adjustments were made.

Other (feature) substitutions, e.g., of the final /θ/ to /t/ in "deaths" /deθs/ to [dets] or of the first /l/ in "tells" /telz/ to /w/ [tewz]. These common substitutions usually involve a simple change of placement and/or manner of articulation (/θ/ → /t/),

while retaining voicing status, or the substitution of a glide (/w/) for an acoustically similar consonant (/l/). They affect intelligibility more than do simple voicing substitutions, as in the following examples: "deaths" → "debts" or "tells" → "tales". What makes them less serious than most deletions is that phonemic data is recoverable once the auditor adjusts to the systematic way a speaker treats a problem consonant such as /θ/ or postvocalic /l/. Other substitutions are given a deduction of 0.3 points. Substitutions to /ʔ/ also deduct 0.3 points, and are similar to deletions in their effects. They merely inform the listener that an extra consonant was intended. No ad hoc adjustments were made.

Unusual substitutions, e.g., of the /z/ in "sings" /sɪŋz/ to /t/ [sɪŋt], or of the /d/ in "allowed" /ʌlaʊd/ to /f/ [ʌlaʊf] are more serious. In the first case the /z/ →[t] double substitution changes both voicing and manner of articulation, while the second, triple substitution changes voicing, manner, and place of articulation of the final consonant. The double substitution changes the apparent grammatical tense of "sings" to an erroneous past tense, while retaining the stem; it has a deduction of 0.4 points. The triple substitution is more serious, for the heard word may be unintelligible; it has a deduction of 0.5 points.[11] These are ad hoc adjustments.

3.5.2.3 Other Errors

Metathesis switches the order of consonants within a word, e.g., the /l/ in "films" /fɪlmz/ to [flɪmz]. In this case the /l/ is taken out of the coda and added to the onset. Since most such examples of metathesis lead to unintelligibility, this error alone has a deduction of 0.5 points.[12] An exception is made for metathesis occurring solely within the coda, e.g., the /sk/ switching in "ask" /æsk/ to [æks]. The Old

English form of "ask" actually used /ks/: 'Acsian' (Schendl, H. 2001: 78), and there are still some regional native speaker dialects which pronounce 'ask' as 'axe'. In any event, there would seem to be a qualitative distinction between metathesis within the coda vs. outside the coda. The former case---/sk/ or /sp/—calls for a deduction of 0.3 points. The evaluator resisted the temptation to make an ad-hoc adjustment to the deduction for coda-onset metathesis downward to 0.4 points, as the token word in question, "films", was generally understood by the auditors because of collocational context, i.e., "watched films"; because [flɪm] is very commonly used by Thai learners to mean "film"; and both auditors are quite familiar with Thai learners. Typical interlocutors exposed to metathesis errors without such context would not fare as well, however.

The final modification to a complex coda involves not deletion, substitution, or metathesis, but vowel addition—either epenthesis between consonants or paragoge after the final consonant. As section 2.3 pointed out, less data is lost with addition than with deletion or substitution. Jenkins (2000) cites a standard Japanese learner's pronunciation of 'product', as [porodukutɑ], where the /o/, /u/, and /ɑ/ vowels are shortened to the approximate length of /ə/ in standard NS speech. This pronunciation turns out to be fairly intelligible in her study, though there are 2 instances of epenthesis and one of paragoge. Jenkins (2000) accounts a Taiwanese speaker's production of [pɒdʌk] as less satisfactory, as the onset has also been simplified.While epenthesis and paragoge are errors, they are often viewed benevolently in the literature (section 2.3.1), and a high epenthesis/deletion ratio is seen as a sign of progress. There was only one speaker in the present study who often used vowel addition.

Informant #3 consistently used epenthesis or paragoge on complex codas derived from **–ed** inflections where the spelling might imply a syllabic allomorph. The following words from the wordlist had epenthesis: *belonged, called, camped, changed, linked, pinched,* and *rubbed.* The following words had paragoge, usually of the front vowel [ɛ]: *asked, fixed, laughed, reached, seemed, sensed,* and *solved. Judged* and *wished* seemed to substitute a glottal stop for the final consonant in the coda, and had effects similar to schwa paragoge. Almost half of these tokens affected by epenthesis or paragoge were adjudged by one auditor to be unintelligible or were assumed to be totally different words, e.g., *called/couldn't, camped/cancel, pinched/pinches, rubbed/rub it, laughed/laughter, seemed/seated, solved/sofier, judged/gesture*. This auditor was even dissatisfied with many of the pronunciations of the words he could correctly make out. The inevitable introduction of an extraneous syllable caused by the additional vowel was in some cases compounded by the introduction of an extraneous consonant as well, e.g., [ɑlaʊdɛt] for *allowed*, [læftɛt] for *laughed*. Epenthesis and paragoge may not be good strategies for this speaker because she is not consistent about using a particular strategy (*rubbed* used epenthesis, while *seemed* used paragoge), and is not careful about limiting her paragoge to a short vowel. Sometimes she seems to use both epenthesis and paragoge simultaneously, e.g., in *laughed* [læftɛt] and *allowed* [ɑlaʊdɛt]. In both cases the paragoge is redundant.

This study would normally weight epenthesis and paragoge errors equivalently to voicing substitution errors, provided that they were implemented in a standard fashion (with short vowels only) and consistently. But epenthesis as used by this informant probably affects intelligibility as much as other substitutions, such as /θ/ →

/t/, and is a bit difficult for the listener who has to deal with an inconsistent alteration strategy. It is given an ad-hoc deduction of 0.3 points, while schwa paragoge without extraneous consonants is given 0.2 points.

3.5.2.4 A Formula for Scoring Multiple Errors

Many, if not most, tokens displayed more than one error in the coda, and the evaluator saw two basic alternative methods of scoring multiple errors: 1) Add up the error points and deduct them from 1.0; 2) multiply the net scores together. The second alternative was provisionally chosen based upon the following common example.

In the case of complex codas with two devoicings, e.g., *used* [juzd] to [just], a deduction of two voicing substitution errors from 1.0 yields a score of 0.6, which is equivalent to a single deletion (0.4), e.g., [juz]. It should be apparent that the grammatical tense is recoverable from the first, devoiced pronunciation [just], but is not in the second example [juz] because the **–ed** phoneme [d] has been deleted. Multiplying, yields (1.0-0.2) X (1.0-0.2) = 0.8 X 0.8 = 0.64. In fact, the interlocutor's ear soon gets accustomed to devoicing, especially when it is systematic. Indeed, in the hundreds of tokens in this study there was only a single example of a voiced complex coda which was pronounced with voicing intact. In every case but one the final consonant of a coda was devoiced even on the rare occasions when the first consonant was voiced. Many, if not most, languages do not have voiced finals, yet many NS soon learn to understand Germans and others who devoice English voiced complex codas, provided the grammatical inflections are still present. In the case of codas with three or more errors, the further ad-hoc step of rounding up the result of the multiplication of the net scores is employed to both simplify and further minimize the

point deduction of such multiple errors. For ease of use, all token scores are reckoned at 100 times the raw (0-1.0) scores. This gives the following outcomes:

Table 2 Illustration of Possible Scoring Outcomes

Points	Error	Example	Calculation
80	**1 voicing substitution**	*boys* /bɔɪz/ → [bɔɪs]	(1-0.2 = 0.8) X 100 = 80
70	**2 feature substitutions**	*deaths* /deθs/ → [dets]	(1-0.3 = 0.7) X 100= 70
	1 coda metathesis	*ask* /æsk/ → [æks]	(1-0.3 = 0.7) X 100= 70
	1 epenthesis	*called* /kɑld/ → [kɑlɛt]	(1-0.3 = 0.7) X 100= 70
64	**2 voicing substitutions**	*used* /juzd/ → [just]	((1-0.2) X (1-0.2) = 0.64) X 100 = 64
60	**1 deletion**	*seemed* /simd/ → [sim]	(1-0.4 = 0.6) X 100 = 60
	1 double substitution	*sings* /sɪŋz/ → [sɪŋt]	(1-0.4 = 0.6) X 100 = 60
56	**1 voicing and 1 feature substitution**	*cows* /kaʊz/ → [kɑ:ʊfs]	((1-0.2) X (1-0.3) = 0.56) X 100 = 56
50	**1 coda-onset metathesis:**	*films* /fɪlmz/ → [flɪmz]	(1-0.5 = 0.5) X 100 = 50
	1 triple substitution	*allowed* /ʌlaʊd/ → [ʌlaʊf]	(1-0.5 = 0.5) X 100 = 50
48	**1 deletion and 1 voicing substitution**	*rubbed* /rʌbd/→ [rʌp]	((1-0.4) X (1-0.2) = 0.48) X 100 = 48
42	**1 deletion and 1 feature substitution**	*deaths*, /deθs/ → [det]	((1-0.4) X (1-0.3) = 0.42) X 100 = 42
	1 deletion and 1 /sk/ metathesis	*asked* /askt/ → [aks]	((1-0.4) X (1-0.3) = 0.42) X 100 = 42
	1 deletion and 2 voicing substitutions	*solves*/sɒlvz/ → [sɒfs]	((1-0.4) X (1-0.2) X (1-0.2) = 0.6 X (0.64, rounded up to 0.7) =0.42) X 100 = 42
36	**2 deletions**	*lamps* /læmps/ → [læm]	((1-0.4) X (1-0.4) = (0.36) X 100 = 36
30	**2 deletions and 1 voicing substitution**	*solved* /sɒlvd/ → [sɒf]	(1-0.4) X ((1-0.4) X (1-0.2)), rounded up to 0.5= 0.3) X 100 = 30

3.5.2.5 Discussion of the Scoring Outcomes

It may seem on the surface that there is very little utility or intelligibility remaining in pronunciations like [sɒf] for "solved" /sɒlvd/, but two things in the rhyme are retained: the vowel and the labio-dental fricative. In context, "We [sɒf] the problem already" may be fairly comprehensible. Similarly, "there were many [det] from the accident", "I [aks] him yesterday", "she [ʌlaʊf] me to go", "I watch some [flɪm]" are certainly decipherable—depending on what else has been said, and the degree of familiarity with the speaker and speaker's dialect. Coda to onset metathesis is probably the most serious single systematic error, yet that it is systematic at all implies that the interlocutor may find the key to learner error. This type of metathesis usually happens when the phonotactic limitations of the speaker inhibit a phonemic sequence, e.g., in "films" /fɪlmz/, "foil" /fɔil/, and "file" /faɪl/, i.e., when /l/ is present as the first consonant in a complex coda or when following a diphthong. In each case, the /l/ is removed from the coda and transferred to the onset, creating a complex onset (/fl/) which is negotiable; we get [flɪmz], [flɔɪ] and [flaɪ], respectively. These are worth 0.5 points. Compounded errors using coda to onset metathesis might involve devoicing of the final consonant [flɪms] (0.4 points) or even deletion of the final consonant with devoicing and other substitution, e.g., in "solves" /sɒlvz/, which was pronounced by informant #2 as [sloʊf]. There are four errors here: 1) coda to onset metathesis (/ol/ → [lo]); 2) deletion of final /z/; 3) devoicing of /v/ → [f]; and 4) other substitution (actually epenthesis) of the /ʊ/. This would be worth ((1-0.5) X (1-0.4) X (1-0.2) X (1-0.3) = 0.188 ~ 0.2) X 100 = 20. This is effectively about the minimum possible numerical score for a pronunciation that is virtually unintelligible. Hence, the

scores might well be normalized from a 20 to 100 scale, with 20 unintelligible and 100 perfect. Had we simply added the various errors in the last example (0.5 + 0.4 + 0.2 + 0.3) and deducted from 1.0, we would have a negative 0.4 X 100 = -40 score for [slouf]. [sɒf] for "solved" would have a score of 1.0-0.4-0.4-0.2, or Zero, yet it was concluded that it might be intelligible in context, e.g., in "We [sɒf] the problem already". This again shows that multiplying the individual net scores together and rounding triple errors up to the next decimal point may be a better option than adding the individual errors together and subtracting from one.

What about adjusting the point deductions upward or downward? If the point deductions were raised (metathesis=0.6, deletion=0.5, other substitutions=0.4, voicing substitutions=0.3) the score for [slouf] would be reduced to 0.112, which still rounds up to the same 0.2 X 100 = 20. The score for [sɒf] would also be 20, which would negate the effective differences between them. The effective differences between deletion and voicing substitution would be reduced. A coda with one deletion (1-0.5) X 100 = 50 would receive about the same score as a coda with two voicing substitutions ((1-0.3) X (1-0.3)) X 100 = 49. Yet, it has been argued that the former error is more serious, as phonemic and grammatical data is lost.

A reduction of the point deductions, say to 0.1 for voicing substitutions, 0.2 for feature substitutions, 0.3 for deletions, and 0.4 for metathesis would yield (1-0.4) X (1-0.3) X (1-0.1) X (1-0.2) X100, rounded up to 30 for [slouf], and (1-0.3) X (1-0.3) X (1-0.1) X100, rounded up to 50 for [sɒf]. This would have the virtue of emphasizing the distinction between truly unintelligible pronunciations (the former) and marginally intelligible ones (the latter). A single deletion would receive (1-0.3) or 0.7 X 100 = 70 points, while two voicing substitutions would receive (1-0.1) X (1-0.1)

=0.81 X 100 = 81 points—which seems about right. The main drawback of a reduction in point deductions would be that one feature substitution (0.2) would be equivalent to two voicing substitutions (2X0.1), and one deletion (0.3) would be equivalent to three (3X0.1), which seems excessive. Thus, the original point deductions (voicing substitutions = 0.2, feature substitutions = 0.3, deletions = 0.4, metathesis = 0.5) appear to be optimal.

Rounding up to the next, higher decimal is employed to minimize the effects of triple errors, to simplify the coda scores, and make them more acceptable in evaluation. Making too fine a distinction by figuring to the thousandth place is not warranted in a scoring system which is intuitive and somewhat arbitrary to begin with. But a comparison of the correlation between coda scores and overall intelligibility outcomes using both raw and rounded up scores might be made to indicate whether rounding up helps or hurts the accuracy gauge. Alternatively, rounding to the *nearest* decimal point might also be tried. These would be further ad hoc adjustments.

A final scoring problem might occur (rarely) when both coda consonants of a doubleton are deleted, e.g., "judged" /dʒʌdʒd/ → [dʒʌ] (score = 36), and this is compared to "solved" [sɒf], above, which had a score of 30. Clearly, the first token would be less intelligible, yet has a higher score. An alternative scoring method which views scores as relating to the proportion of the coda remaining intact might be employed. The first token here would have 100% of the coda deleted, while the second token retains one third of its original tripleton coda--if de-voiced. As there was actually only one token with a doubleton coda in the present evaluation which was totally deleted ('called') [kɔ:], this alternative scoring method was not pursued.

3.5.2.6 Other Ad Hoc Scoring of Codas

Section 3.5.2.4 gave several examples of ad hoc departures from the broad speaker-centered scoring metric (voicing substitutions, 0.2; feature substitutions, 0.3, deletions, 0.4; Coda-onset metathesis, 0.5). Arguments have been made for ad hoc adjustments to some medial consonant deletions, unusual double-substitutions (0.4), triple-substitutions (0.5), coda metathesis (0.3), and epenthesis (0.3). This last error may generate many additional complex situations which call for ad hoc adjustments. The following are a dozen actual tokens from the wordlist with their scorings and rationales for scorings.

1. [bilɒŋkt] for *belonged* (80). The intrusion of /k/ did not affect intelligibility. One voicing substitution with a 0.2 deduction yields (1-0.2) X 100 = 80.

2. [lʌmf] for *lamps* (42). The /f/ is taken as a feature substitution either for /p/ or for /s/, leaving one deletion, either of /p/ or /s/. (1-0.3) X (1-.04) = 0.42 X 100 = 42.

3. [tæksk] for *tasks* (60). A double coda metathesis is apparent: /sks/ → /ksk/. The word ended up sounding like 'task', with deleted final /s/. (1-0.4) X 100 = 60.

4. [entʃɔɪn] for *enjoyed* (70). The /d/ → /n/ substitution may seem odd, but /d/ and /n/ are both voiced and share the same place of articulation; thus, a 0.3 feature substitution deduction was taken.

5. [lɪpf] for *lips* (70). The /s/ →/f/ substitution may also seem odd, but is actually rather common. Two voiceless anterior fricatives warrant a feature substitution. (1-0.3) X 100 = 70.

6. [pɪntʃʔs] for *pinched* (60). An unusual double substitution which erroneously indicated present tense. The stem is still recognizable. (1-0.4) X 100 = 60.

7. [tæks] for *tasks* (60). Although coda metathesis is probable here, it seems simpler to just count the deletion of the first consonant, /s/. (1-0.4) X 100 = 60.

8. [drækdəʔ] for *dragged* (64). The /əʔ/ had a similar effect to schwa paragoge (0.2 point deduction); with one voicing substitution, we get (1-0.2) X (1-0.2) X100=64.

9. [tʃʌts] for *judged* (42). The /ts/ was taken as a feature substitution for /dʒ/ (0.3 points). The final /d/ was deleted. (1-0.3) X (1-0.4) X 100 = 42.

10. [fɪksdə] for *fixed* (70). Schwa paragoge would normally rate a 0.2 point deduction; however, the final /t/ was erroneously voiced, so 0.3 was deducted. (1-0.3) X100 = 70.

11. [kɒlɛt] for *called* (70). Epenthesis using a long vowel rated a 0.3 deduction.

12. [rʌppɛt] for *rubbed* (56). In addition to epenthesis, the /b/ is devoiced. (1-0.3) X (1-0.2) X 100 = 56.

From the listener's point of view, those errors which have been experienced most often may be the easiest to interpret, and it is possible that some words involving deletions might be easier to decipher than some words involving substitutions, especially uncommon ones, like some feature substitutions. Certainly, many deletions of the medial consonant in a triple consonant cluster may be expected, and final consonants may undergo elision in connected speech. But words spoken individually in isolation should maintain the inflections, or they are simply deficient, and indicate that the speaker will very likely consistently delete a given coda. Thus, the wordlist tokens are an important measure of the learner's capability.

3.6 Subsequent Student Evaluations

The evaluator performed subsequent evaluations on a group of eight less advanced students (hereafter referred to as "students", not "informants") who are preparing to enter the same international program in which the four informants of the main study are already enrolled. Five of these eight students are Thai nationals; one is Chinese, one, Korean, and one, Japanese. The evaluator has been teaching these students elements of oral communication, and has trained them in rudimentary English phonetics, as their oral skills are much weaker than their reading and writing skills. After one month of such instruction the author recorded them reading aloud the first paragraph from the Bird Flu narrative into a portable Sony micro-cassette player. Audio quality was not as good as a notebook computer and evaluation could not make use of visual audio wave forms, as provided by the SIL 2.2 Speech Analyzer program used in the informant evaluation. Nevertheless, a micro-cassette recorder may be all a typical teacher has available as a classroom resource.

Because less phonetic detail was obtainable from less accurate equipment, evaluation involved less detailed scoring of the coda pronunciations than in the main study. Hence, the scores in this evaluation are not strictly commensurate with the scores in the informant study. A simple 10-point scoring scheme was used, and the results of multiple errors were rounded up to the next point. Deletions were 4 points; feature substitutions and epenthesis, 3; voicing substitutions and schwa paragoge, 2, and metathesis, 5 points. Two common complex codas (/nt/, /sk/) which were not evaluated in the main study were assessed. As with the informants, the students were given very little preparation before speaking, but were told to pronounce the inflections clearly. They were able to pronounce most of the codas, though they had

particular difficulty with some voiced **–ed** codas. The scoring took about three hours for 240 tokens (80 tokens per hour) as compared with perhaps 5 times this amount using the SIL program in the first (informant) evaluation and about 15 times that amount through final assessment and tabulation. The student scoring was more lenient, as well as simpler than the informant scoring: some devoicing substitutions and deletions did not receive full deductions if they seemed fairly clear to the evaluator. Thus, the assessment was more subjective as well.

.

3.7 Summary

Limiting evaluation to the pronunciation of the complex coda rather than of entire word-tokens is more realistic for a study such as the present one. A binary pass/fail assessment is rejected in the study because it does not make useful distinctions between the various error types, and it assumes that all errors are equivalent in gravity and would lead to equivalent intelligibility outcomes. Instead, errors which are generally considered to be both more grave (in that they reduce or mar phonemic data which can be conveyed) or which reflect earlier stages of complex coda acquisition were assigned more weight in deduction than were other errors which do not mar phonemic data to the same extent and which are characteristic of more advanced learners. Error types are tabulated for all codas, and used to generate numerical scores so that the codas may be ranked by difficulty. In order to validate the scoring method used in the study, individual coda scores are correlated with overall intelligibility of the informants who utter the codas. Finally, since it is TESOL instructors who must ultimately assess and effect improvements, the test and scoring method are checked for reliability with a single evaluator and different group of learners using standard equipment available to most teachers.