The Production of a More Detailed Description of Problems Experienced by Native Japanese Speakers When Articulating Certain English Speech Sounds (2) : Error Analysis

Timothy Ashton*

Abstract: This paper is a direct continuation of the study initiated in Ashton (2006) in which an English-Japanese Contrastive Analysis (CA) was conducted of twenty selected English speech sounds and sequences (the consonants [ f, v, ð, œ ] in various syllabic positions and sequences of the consonants [ w, j, tʰ ] with various vowels) and the Japanese speech sounds that most closely correspond in articulatory terms and common conditions of occurrence. The aims of this study were to investigate the degree of difficulty surrounding each sound in these syllabic conditions and use the results to supplement information in two selected published texts, Joanne Kenworthy's "Teaching English Pronunciation" (first published 1987; Longman) and "Pronunciation Contrasts in English" by D. L. F. Nilsen and A. P. Nilsen (1971; Regents Publishing Co.), which have been widely used to implement regular English pronunciation practice to classes of Japanese learners. Predictions of error on the part of native Japanese learners were

* Foreign Language Instructor, Fukuoka University.
posed, based on contrasts in the phonetic descriptions of the English and Japanese sounds. However, from the CA alone it was impossible to address all of the aims of the study, leaving it necessary to test the validity of the CA-generated hypotheses by means of an Error Analysis (EA) survey, the results and conclusions of which could be used to address the issues and aims left inconclusive by the CA.

The EA survey consisted of a two-part diagnostic test, which was designed to investigate the aural and oral phonetic skills of a group of twenty native Japanese learners, with regard to the above English speech sounds. The linguistic backgrounds of the learners were surveyed by means of a personal questionnaire. Safeguards were incorporated into the EA methodology to ensure that each speech sound was tested orally and aurally at least once, by reliable and valid means. Additional speech samples of each sound were also found within the testees’ oral responses to cross-check quality of articulation.

A number of errors were observed that were both predicted and not predicted by the CA. Nonetheless, they were mostly hypothesised as the result of L1 Interference. It was concluded however, that the CA could have also predicted these errors had it contained more detail. Finally, the combined findings of this study were then worded as supplementary comments to references already made to the above sounds in the aforementioned texts, with subsequent relevance to the direction of possible EFL materials. The implications of these proposals to EFL teaching were also considered.
1. Introduction: A Review of CA/EA Literature

The late 1950s through to the early 1970s saw seminal works by (among others) Lado (1957) and DiPietro (1971) advocating CA theoretical and methodological frameworks and their applications (e.g. those mentioned in 5.2). In the 1970s authors (such as those in Richards’ 1974 collection) seemingly damned CA, electing EA as its natural successor. Interlanguage theories by Selinker, Nemser, Corder et al were also published. Works in the 1980s have since concluded that CA and EA, when used in collaboration, produce more comprehensive results. Recent literature seems to focus on two directions of research. One of these directions appears to be the redefinition of ‘old’ CA/EA concepts and theories. Examples of these are James (1994) who reviews further distinctions made in Transfer theory, also calling for redefinition of other terms, and Hammerley (1982) who redefines Stockwell and Bowen’s ‘phonological hierarchy of incremental difficulty’. Such works, although interesting, do not seem to shed much further light on existing notions but merely increase the taxonomy already involved.

The other direction of research appears to be of greater interest. This consists of recent accounts of ‘new’ CA/EA-based research; more ground-breaking work challenging older hypotheses and generating new ones. With regard to English-Japanese studies, the pages of numerous journals (e.g. "IRAL", "Language Learning" and "Language and Speech") have seen a shift in focus to Acoustic Phonetics over the past decade. In particular, great interest has been shown in the perception, rather than the production, of English [l] and [4] by Japanese learners, e.g. Henly and Sheldon
(1986), Flege et al (1995) and Miura (1996). CA/EA speech production studies still flourish however, e.g. Lenhardtouá (1993) of the English of Slovaks. Despite criticism over the decades CA/EA-based research has maintained its momentum. There can be little doubt that CA and EA still have their places in contemporary linguistic research.

2. Predicting and Explaining Problems in L2 Learning: The Study of Learners' Intermediary Linguistic Systems

Various terms have been proposed as labels for the phenomenon of intermediary linguistic systems, e.g. 'Interlanguage' (IL). Nemser (1971: 55) introduced the term 'Approximative System' which he defined as "the deviant linguistic system actually employed by the learner attempting to utilise the target language". This also implied its developmental nature through continual modification from new elements incorporated throughout the learning process (Richards 1971: 29-30). Corder (1981: 14-18, 68, 90) comments on the term 'Idiosyncratic dialect', emphasising its unstable status, as the learner attempts to bring his communicative behaviour in line with that of the target language, group or community. He also suggests the alternative label 'Transitional Dialect' acknowledging its intermediate status (between the L1 and L2). Corder also points out that this dialect should be regarded as a continuum with its own developmental history rather than a hybrid language in itself. The most appropriate term however (and the one that will be used throughout this study) appears to be Selinker's 'Interlanguage' as it encompasses all of the above, suggesting (according to Sridhar 1981: 227):
The intermediate state of the learner's system.

Its instability, including the rapidity with which it changes.

Recognition of a rule-governed, often adequate communication system, i.e. a language.

3. Intralingual and Developmental Psychological Processes Resulting In L2 Performance Errors

Selinker (1972) suggests the existence of a latent psychological structure in the brain which is activated whenever attempts are made to produce meanings. This structure putatively contains five central processes which control to a very large extent the surface structures of IL utterances (Selinker 1972: 37, 49, Richards 1971: 29, Sridhar 1981: 229). These five central processes are:

(i) Language Transfer: L1 habits predominate in the performance of the L2 until new language habits can be acquired (Corder 1981: 5). Along with grammatical, lexical, syntactic and morphological (etc.) systems, the individual tends to transfer the whole L1 sound system, including stress rhythm and intonation patterns (Lado 1957: 11-24).

(ii) Transfer of Training: positive transfer (facilitation, due to similarities) and negative transfer (interference, due to differences) of language training procedures onto other skills. Van Els et al (1977: 49-50), makes the further distinction of 'pro-active transfer' (of existing skills onto new skills) and 'retro-active transfer' (of new skills onto existing skills). For example, Japanese learners may produce all question-sentences with a rising intonation pattern due to the sheer frequency that such patterns are
presented in high school textbooks and drilled by teachers.

(iii) **Strategies of Second Language Learning:** identifiable approaches by the learner to the material to be learned. For example, the positive transfer of Japanese double vowels to aid the production of English diphthongs, e.g. [əː] as in 'gate' may be substituted with Japanese [ei] as in 'eiga' (えいが: 'movie, film') as an initial step in the mastery of [ɛi].

(iv) **Strategies of Second Language Communication:** identifiable approaches by the learner to communication with (native and other) speakers of the target language. For example, a learner (Japanese or otherwise) may consciously disregard the correct production of many problematic areas of pronunciation for fear of his/her speech sounding unnatural or exaggerated (Selinker 1972: 40).

(v) **Overgeneralisation of target language material:** clear overgeneralisation of target language rules, semantic features, etc. For example, a learner may assume that all syllables containing a vowel sound, but that do not carry main/primary stress, involve a reduction of the vowel to 'schwa' [ə] (which is not the case in words such as 'heating' [ɪ] or 'constitution' [ʊ, ɪ]'.

Selinker’s claims hardly seem revolutionary; of course linguistic processing activity occurs in the brain (!) and so the description of a 'latent psychological structure' is not particularly explanatory. The above notions are also not the product of observation and so could even be regarded as mere 'armchair speculation'. Selinker goes on to claim that in addition to these five 'central processes' there are also many minor ones that should not be

---

1 ‘= syllable carrying main or primary stress within a word.
ignored, for example:

a) **Spelling pronunciations:** e.g. pronunciation of the word 'quay' as [kwai].

b) **Cognate pronunciations:** e.g. pronunciation of the word 'student' with an additional final vowel [ə] ('studento') by many Japanese learners regardless of whether they are capable of avoiding this or not.

c) **Holophrase learning:** e.g. the production of 'one and half-an-hour' from the learning of 'half-an-hour'.

d) **Hypercorrection:** erroneous attempts at correct usage due to inadequate knowledge of conventions e.g. in trying to produce an aspirated [tʰ] a Japanese learner produces a sound resembling [θ] to the English ear.² (Kenworthy 1987: 150)

Richards (1971: 174-181) extends Selinker's fifth process, positing the existence of 'Intralingual' and 'Developmental' errors. Intralingual errors are "those which reflect the general characteristics of rule learning." Developmental errors "illustrate the learner attempting to build up hypotheses about the L2 from limited experience of it in the classroom or textbook" (although it could be argued that such a distinction is unnecessary and that a general heading of 'Intralingual errors' would suffice). In addition to 'overgeneralisation' he also adds the following as causes of error:

(i) **Ignorance of rule restrictions:** the application of rules to contexts where they do not apply. For example, a learner may apply the rules of final and initial consonant elision, e.g. 'jam making, good dog' to circumstances in

² Not to be confused with the alleged transfer of the Japanese sequence [tʃi] to English words beginning with [ti] e.g. 'team'.
which they do not occur such as $[dʒ] + [dʒ]$ ('large gin'), $[v] + [v]$ ('rich journalist')$, [v] + [v]$ ('Dutch cheese') etc.

(ii) **Incomplete application of rules:** deviances that represent the degree of development of rules required to produce acceptable utterances. For example, a learner may understand that usually items such as articles or prepositions do not carry any stress in a sentence but that they can in a contrastive context.

(iii) **The building of false systems or concepts:** developmental errors that derive from faulty comprehension of distinctions in the target language. For example, after learning that sounds such as $[f, v], [s, z], [θ, δ]$ are voiceless and voiced (respectively) counterparts of each other, the learner may go on to hypothesise that somehow all consonant sounds have a voiceless or voiced 'equivalent' (e.g. a voiceless sound of similar phonetic grouping to $[m]$ or $[n]$ etc.)

Richards (1971: 182, Richards and Sampson 1974: 6) describes the above categories of errors as 'non-contrastive' in origin, i.e. not due to L1 transfer. Instead they are derived from language learning strategies employed by the learner, the transfer of training and the mutual interference of items within the target language. Errors of this nature are frequent and irrespective of the learner's background. One should note, however, that it is extremely difficult to unambiguously identify which processes observable errors in L2 are to be attributed to, as is often the case with psychological notions and observable behaviour (Selinker 1972: 42).

The above theories will be related in an explanatory capacity (where possible) to the results of the EA survey conducted during the course of this study.
4. Fossilisation

Very often the learner's IL reaches a state adequate for communication (with either native speakers or as a lingua franca) and motivation to improve wanes. At this stage, IL rules and subsystems perceived as erroneous in terms of the L2, become 'fossilised', re-emerging in L2 performance even when thought eradicated. Future progress or improvement in the L2 then seems unlikely. Similar to this is Nemser's notion of 'Plateauing', described as a stable variety of IL (Selinker 1972: 36-37, Corder 1981: 87-88). This can be found, for example, in groups or communities of immigrant workers, where it has even been regarded as a communication strategy. Martinet (1962: 138) has also theorised about a 'law of minimum effort': the desire to achieve ends with the least effort and how it can conflict with communicational clarity. In many cases, the gain in communicational clarity is less than the effort needed to change.

Nemser accounts for fossilisation as the results of two forces: 'demands of communication' force the establishment of phonological, grammatical and lexical categories in IL. 'demands of economy' force a balance or order in the linguistic system. Both are restricted by the capabilities of the learner at that stage of development. Selinker (1972: 49) however, points out that the 'latent psychological structure' in the brain (see 3) also contains a 'fossilisation mechanism'. This forces fossilised material onto IL utterances. Unfortunately, due to the constructs of the brain being not directly observable, these claims remain merely speculative as do the interlingual and intralingual/developmental concepts described earlier. The only
'evidence' lies in learners' L2 performance.\(^3\)

5. An Error Analysis Survey of the Aural and Oral Phonetic Skills of a Group of Native Japanese Learners of English

This section reports the design, implementation and findings of an EA survey, motivated by the hypotheses drawn from the English-Japanese CA described in Ashton 2006. The aims of this survey were:

i. To test empirically the descriptive data generated by the CA (see Ashton 2006: 37-41)

ii. To provide further empirical data concerning learners' actual performance errors for additional analysis based on the intralingual / developmental psychological theories outlined in 3.

5.1 Terminology

The following terms appear in this section:

'Test reliability': "Concerned with ensuring that a test is an accurate measure" (Davies 1990: 6). A test may be regarded as 'reliable' if it meets the following criteria:

(i) The scores generated by the test do not fluctuate considerably (or for no apparent reason) if used on the same candidates repeatedly, provided that

\(^3\) A motivational analysis was not carried out prior to this study and so it was not determined whether the test group wanted to improve or not. Fossilisation will therefore be considered a variable when the results of the EA survey are analysed (see 6.5).
there has been no intervening instruction.


'Test validity': "Attempts to provide a theoretical framework which gives reassurance to the test" (Davies 1990: 6). It is concerned with the linguistic content of the test and the situation or technique used to test this content, i.e. the test must be meaningful and appropriate. A 'valid' test must satisfy the following criteria:

(i) The test measures what it is intended to measure and nothing else.

Face validity is concerned with what testers and testees think of the test. It can be determined by both parties simply by inspecting the test and giving their opinions on it. The test involved in this study is too brief for face validity to have significant objective value, however.

Test reliability and validity are not coterminous considerations but complementary aspects of a common concern in measurement. A test also cannot be valid unless it is reliable; a means of measurement that it is inconsistent, thus producing inconsistent scores, an essential ingredient to general validity (Bachman 1990: 25, 160, Heaton 1975: X, Underhill 1987: 105).

'Objective scoring': involving a single correct answer unambiguously identified as either 'correct' or 'incorrect'. No evaluative judgement is required of the marker and the test may be marked mechanically.

'Subjective scoring': requiring the subjective judgement of the marker.
'Objective' and 'subjective' are terms referring to the scoring system only and not the format of the test. All tests and test items are initially constructed subjectively by the tester/examiner (Underhill 1987: 8, Lado 1961: 28-29, Harrison 1983: 110, Hughes 1989: 19, Heaton 1975: 11).

5.2 A Methodological Framework for the EA Survey

The following four-stage framework was used throughout this survey. It involves the design and implementation of an EA test, the marking and grading of the testees' linguistic responses and theoretical explanation of the errors observed in the L2 performance of the errors involved. The framework is based on methodologies reported by Sridhar (1981: 222) and Van Els et al (1977: 47). These frameworks have been referred to as they are claimed to outline EA methodology in its most generalised terms, as followed by the majority of EA practitioners. The framework below differs from those published by the above authors however, in that some stages have been omitted due to the restricted nature of this study. Additional stages may involve, for example, determination of the degree of disturbance caused by the error in terms of its effects on communication, and suggestions for learner remedial work.

1. Collection of data: The design of an EA test to specified criteria, its implementation and consideration of test variables with subsequent effects on the reliability and validity of the testing procedure.

2. Identification of errors: Identification of the learners' failure to adequately perceive and/or produce given speech sounds. This stage of the framework first covers the design of a grading system, and then its
implementation. The effects of the grading system on the reliability and validity of the testing process are also considered.

3. **Statement of findings:** The results of the EA survey in terms of the frequency of correct or incorrect aural responses or oral articulations of test items involving the speech sounds under test. Such findings may be expressed in percentages.

4. **Description and explanation of errors:** Theoretical analysis (e.g. interlingual or intralingual) of the source of the errors observed, inferred from the learners' L2 performance.

5.3 **Collection of Data: Design and Implementation of the EA Test**

5.3.1 **The Importance of Testing both Oral and Aural Phonetic Skills**

Recognition and production of L2 items are separate sub-skills and must therefore be tested separately using different techniques. Any IL study should be based on the testing of both sub-skills in order to effectively ascertain the degree of development of an individual’s L2 phonetic skills. Lado (1961: 59, 71, 78) and Valette (1967: 87) offer theories as to why this is so:

Aural skills are frequently more advanced than oral skills, i.e. learners are often able to perceive L2 sounds and sound contrasts before being able to produce them. This is possibly because production generally requires more knowledge and mental processing than perception. By the same token, learning to identify L2 sounds does not imply the ability to produce them. In addition, the differences in ability between oral and aural skills vary from learner to learner. Some learn to discriminate quite well but are
unable to effectively articulate, others learn to articulate as well as they discriminate. Consequently, assuming that aural skills are more advanced than oral skills, oral communication problems are more likely to be experienced by the learner than aural. Conceivably, a learner could fail to adequately produce a desired word (due to a limited articulatory range) but recognise it straight away when listening. It could therefore be concluded that if only production were tested, it would not be ascertained what the learners were able to perceive and vice versa. Verbal behaviour in any case involves acting as the sender of a message as well as the receiver. To identify overall phonetic skills both production and reception must be tested.

5.3.2 Test Type

This study involved a two-part diagnostic test, very similar in design to that described in Ashton 2004: 5-11. It was designed in order to expose specific strengths and weaknesses within the individual learner’s oral and aural phonetic skills, with regard to the twenty English speech sounds and sound sequences listed and examined in the CA (Ashton 2006). These were:

i. [ f, v, ð, t ] in initial, medial and final positions, including when orthographic ‘r’ is represented as [ə] in syllable final.

ii. [ w ] followed by [ i, əi, v, i, ɔ ]

iii. [ tʰ ] followed by [ i, uw ]

iv. [ s ] followed by [ ə, əi ]

The data produced was tabulated both as an overall score for each learner and as lists of the test group’s overall strengths and weaknesses concerning
these sounds: see fig. 1-8 (Hughes 1989: 13-14, Underhill 1987: 13). Bachman (1990: 60) notes that diagnostic tests can be either theory or syllable-based. This test was theory-based, motivated by the hypotheses drawn by the preceding CA and the psychological notions of language learning and communication as described in 3 and 4.

Part I of the test examined aural skills and consisted of a speech sound discrimination exercise as suggested by Lado (1961: 369) and Heaton (1975: 165). The exercise utilised Nilsen and Nilsen's system of contrastive pairs of sounds and utterances (see Ashton 2006: 7-8) and was based on the premise that the ability to discriminate the sounds of a language may initially be equated with the ability to distinguish contrastive pairs of sounds (Valette 1967: 48). Later on in L2 acquisition, the learner may rely on lexical, grammatical and contextual clues to help discriminate troublesome sounds, which can confuse the meanings of utterances. Before such progress can be made however, the learner may depend more upon knowledge of basic sound contrasts within the L2. Considering the testees' presumed state of L2 development (i.e. not that of so-called 'beginners' but also not of 'advanced' learners) it could be reasonably assumed that the members of the

---

4 Such exercises are also referred to as 'phoneme discrimination tests' by many authors (e.g. Lado 1961) and researchers. Such terminology is, however, inconsistent with the definition of a 'phoneme' in which a phoneme is viewed as a model (a phoneme corresponds to an infinite class of actual articulations (speech sounds) via its allophones, speech sound types with the same function in communication. If used in the above context it is (a) not isomorphic with speech behaviour (b) already distinguished and (c) it assumes that speakers somehow 'have' or 'use' phonemes. The term 'speech sound discrimination test' involves less theoretical presuppositions and is therefore more appropriate.

5 Discrimination is only one of three areas within the skill of listening that require proficiency. The other two are retention and comprehension, untested here due to the restraints of the present study (Valette 1967: 48).
test group would be more inclined to rely on sound than lexical or grammatical (etc.) cues to determine meaning.\textsuperscript{6}

Part II of the test was designed to test the learners' oral skills. Visual stimuli (pictures) were used to quickly and clearly convey the meanings of the 'test words' containing the particular speech sound or sound sequence to be elicited. Pictures were used on the assumption that the testee would be less able to spontaneously produce a sentence containing a word (and subsequently test speech sound) that he or she does not understand. (Lado 1961: 82, Valette 1967: 91, Heaton 1975: 90, Underhill 1987: 66-67).

\textbf{5.3.3 Description of Testing Procedures: Part I (Aural)}

The testees listened to a recording of 20 sentences on a cassette. These sentences were also written on their answer sheets but with a 'test word' containing the 'test speech sound' paired up against a similar word, which had another sound in direct contrast with the test sound. These two words were isolated from the rest of the sentence in brackets, with an empty box next to each. The testees ticked the box next to the word that they thought they heard on the tape. Bachman (1990: 129) citing Popham (1973) refers to this as the 'selected' response; the identification of a correct (or incorrect) answer. An example is given overleaf.

The sound contrasts distinguishing the words in each pair affected the meanings of the words and consequently the meaning of the sentence. Both words in each pair were plausibly correct within the context of the

\textsuperscript{6} These notions are purely speculative. Determination of which system of cues the testees relied more upon would require a separate investigation prior to this study.
sentence. The testees were given no grammatical, lexical or contextual clues as to the correct answers. All responses were therefore based purely on aural discrimination.

The tape was played only once (a total playing time of approximately three minutes) with each sentence uttered only once; the testee could have had a better chance of choosing the correct answer if allowed to benefit from the learning effect of hearing each item a second time. The voice on the tape was that of this researcher, a speaker of Standard Southern Spoken British English (thus the English dialect selected for this study was controlled). Care was taken to deliver each sentence in a fashion that was clear and natural in pace but in which the natural stress-rhythm of the sentence and articulation of the test speech sound were unexaggerated. No indication was made to the testee as to which sound was being tested in each item (as this would have given away the answer) although of course the pairs of contrasting sounds could easily be identified. See Appendix II for a sample of Part I of the test.

**Example:** Q.15 tests the recognition of the sound sequence [tʰi] (the correct answer) against the sequence [θi]:

15. What a good □ theme. □ team.

**Part II (Oral):**

The testee was confronted with a series of pictures on a sheet of paper with ‘test words’ containing the ‘test speech sounds’ written underneath. The pictures were used to provide the testee with a visual prompt to form
sentences containing the test words and subsequently the test speech sounds. The test items followed the general theme of 'describing your teacher'.

The same twenty speech sounds and sequences were tested as in Part I. There were twenty individual pictures with twenty corresponding test words. Each testee spontaneously produced twenty sentences that were recorded onto a digital mini-disc for grading later. Bachman (1990: 129) citing Popham (1973) refers to this as the 'constructed response', in which the language sample is produced in response to input material. This part of the test took no longer than five minutes per testee. As in Part I, no indication was made to the testee as to which sounds were being tested in each item as the testee would then have had the opportunity to focus on the articulation of the test sound. This could have resulted in an exaggerated response. See Appendix III for a sample of Part II of the test.

**Example:** the second item from Part II of the test elicits the sound sequence \[ j_\& ] as in 'shave':

Possible testee response:
"Tim *shaves* every day"


5.3.4 Test Design Criteria

Errors are, by definition, systematic and therefore require systematic procedures of elicitation (Lado 1961: 80). Each speech sound or sound
sequence has to be tested at least once under manageable, reliable and valid conditions, in which the testee is provoked into making a linguistic response based upon the limits of his or her IL. Where possible, the test sounds should also be cross checked with other samples of the same sound generated within the boundaries of the same test and under the same conditions. A high degree of systematicity, reliability and validity can be achieved through adherence to the following criteria in the design of this particular EA test:

i. Formal testing.

ii. Discrete point/item testing.

iii. Direct testing.

(i) Formal Testing: Although both parts of the test were conducted in the relatively informal setting of the testees' EFL classroom, a formal approach to the test needed to be adopted. The testees' oral and aural phonetic skills were examined by means of structured, scripted exercises (Part I: speech sound discrimination, Part II: visual stimuli) carried out by the test group for an examiner within pre-determined time restrictions.

(ii) Discrete Point Testing: Hughes (1989: 16) defines discrete point testing as the testing of one element at a time, item by item. In both parts of the test the testee's attention is drawn to specific words (or in the case of Part I, specific pairs of words) containing the sounds or sequences under test. For example, in Part I, Q1 tests the sequence [ wɔə ] as in 'warder', Q2 medial [ v ] as in 'rover', etc. In Part II, initial [ f ] as in 'funny' is tested, then the sequence [ fɛi ] as in 'shave', etc.

(iii) Direct Testing: According to Hughes (1989: 15) direct testing can be
defined as "when it requires the candidate to perform precisely the skill we
wish to measure". The test involved in this study was inherently direct as
it was an examination of observable verbal behaviour. Hughes goes on to
mention that direct testing has the disadvantage of being limited in its
range of task and test items. However, as the aim of this test was to
examine only a predetermined number of speech sounds and sequences, the
use of direct testing was not considered a particularly significant disadvan-
tage.

Free speech (reflective of indirect testing), either recorded for the Part I
aural test or produced by the testee as raw data in Part II, would not have
been a particularly reliable, valid or manageable option. Too much data
could easily have been generated and relevant data still missing. Instead
lexical, grammatical and contextual (etc.) restraints needed to be imple-
mented in order to guide the testee into making specific choices (Part I) and
producing specific utterances (Part II). Such restraints are not so unrealis-
tic or artificial as they may appear however, as any stretch of speaking is
constrained by the situation as well as the other speaker. For example, both
of the test words used in Part I, Q15 were nouns (‘theme/team’; see test
example, 5.3.3). No further grammatical clues were provided and both
words are semantically acceptable within the given context so the testee's
response was restricted to aural discrimination. 'Shave' (the second test
item in Part II) could be either a noun or a verb, but in either case the testee
was obliged to produce a sentence in which the test word was lexically,
grammatically and contextually acceptable.

The test was not entirely direct, however. For it to have been purely direct
it would, strictly speaking, have to have involved the perception and

(20)
articulation of the test sounds in isolation. In both Parts I and II the test sounds were partially contextualised within words and sentences. The setting for each sound was therefore not totally artificial but reflected the need for a balance to achieve a reasonable size of phonetic, grammatical and semantic control.

However, no linguistic linking devices between the sentences were provided in Part I or encouraged in production in Part II as would perhaps be the case in authentic speech. Such detail seemed unnecessary as the focus was only on a single sound in each case. In Part II, the testees' sentences were linked by a common theme ('describe your teacher'; see 5.3.3) but only as a contextual aid to hasten production. Even this was unnecessary in Part I as the exercise was based purely on aural discrimination of sounds. This test should therefore be described as a 'partially contextualised direct test'.

5.3.5 Intrinsic and Extrinsic Factors Affecting Test Reliability and Validity

A number of intrinsic and extrinsic factors can affect the overall accuracy of a test that are not directly associated with the testee's language ability. Such factors include (for example):

- Failures of the test itself.
- The way it is scored.
- Circumstances in which the test is taken.
- Personal factors of the individual testees (health, motivation, etc.)

In order to answer the simple question "Does the test work?" in the sense that it maximises the effects of the language under measurement, such
factors must be considered in terms of their effects on the test's reliability and validity (Bachman 1990: 160-161, Underhill 1987: 105).

(i) Guessing of lexical items: As mentioned in 5.3.4 (iii), there were various lexical, grammatical and semantic (etc.) constraining factors, as well as the advantages and disadvantages of each, that needed to be considered and applied to the lexical items (containing the test sounds) involved in the test. Such restraints have the advantage of allowing the hypotheses motivating the test to be concentrated on but suffer the disadvantage of forcing the inclusion of items unknown to some or all of the testees. This, in turn, could encourage wild guessing which would reduce test reliability. In Part I, as with all two-choice tests, there was a 50% chance that each answer could have been guessed correctly. In Part II, the guessing of an articulation could have potentially skewed the results. The following guidelines needed to be followed when selecting lexical items for inclusion in this test:

For Part I (aural), pairs of the test words needed to be selected that were compatible as parts of speech (so that the testee could not deduce the correct answer from the grammatical or syntactic context) and which followed the same syllable stress patterns. For example:

Q.1: 'warder/hoarder' are both nouns with their main/primary stress falling on the first syllable.

Q.9: 'funny/honey' are adjective and noun respectively but are syntactically acceptable in the same place in the test sentence. They also carry main stress on their first syllables.

For Part II, lexical items that were non-ambiguous, picturable and with only one correct, current pronunciation needed to be selected. Test words containing long sequences of three-four consonants were also to be avoided
as the orthography involved may have appeared daunting to the testees and caused them to unconfidently guess articulation. Test words such as 'thief' (to test final [f]) and 'wood' (to test the sequence [wʊ]) were therefore selected as they do not contain such factors.

Test words should also fall within the testees' active vocabulary range. Selection of such items may be based on teachers' intuition of lexis expected at this stage of IL development. Where this seems less possible, the meanings of potentially unknown items can be explained by the examiner (also upon the request of the testees) before implementation of the test.

(ii) Use of tape recorders / digital recording devices: The test group of twenty candidates was too big to fit into a single classroom for administration of Part I of the test, and so was divided into two groups of ten. These two groups were then tested in direct succession with no contact allowed between the groups and under the same test conditions. This situation therefore demanded uniform spoken delivery of the test items. This was achieved through use of a tape recording of the test sentences ensuring identical delivery on both occasions. This process carried high reliability. For Part II, the voices of the testees were recorded onto digital mini-disc which enhanced scoring and inter-marker reliability. The testees' responses could be played as many times as deemed necessary to properly assess the production of any given speech sound. Attention was also paid to the quality of the recording for both parts of the test as poor quality reduces

---

For Part I testing particularly, it was very difficult to find pairs of words that were: (i) compatible as parts of speech, (ii) followed the same stress patterns, (iii) different only in the articulation of a single sound or sequence, (iv) both semantically possible, and (v) all within the testees' active vocabulary range.
validity. Recorded speech only results in a certain degree of validity however, as any recording distorts the voice to some extent and a disembodied voice does not represent a realistic situation, which would be visually aided by gestures and facial expressions. Higher reliability could have been achieved through use of video tape (Lado 1961: 47, Heaton 1975: 58, Valette 1967: 23, Underhill 1987: 42).

(iii) Testees' sense of familiarity with the test conditions: Reliability and validity can be further enhanced if the test group share a high and equal amount of familiarity and ease with certain aspects of the testing process: During all phases of the current test, the testing environment consisted of the same, quiet classroom with adequate temperature, acoustics, suitable seating and lighting. All testing was administered outside of scheduled lesson hours to lessen the chances of distracting sounds or movements. The candidates' teacher (this researcher was not their teacher) would have made a valid tester; the teacher being a real part of the candidates' lives, tensions and artificialities would be reduced. By the same token however, the teacher/tester could also be a potential danger to validity (see 5.4.2; 'sympathetic response').

The wording and 'metalanguage' of the rubric needs to be within the testees' lexicon as this can affect the validity of the rubric itself. Unless testees can follow the instructions given, the test itself will be neither reliable nor valid (Hughes 1989: 39, 106, Heaton 1975: 98, 160-161, Underhill 1987: 40).

Explanation of the rubric was demonstrated by the tester to the test group (for Part I) and each individual testee (for Part II) in a uniform manner before implementation of the tests.
The learners involved in pilot testing (see 5.3.7) were not included in the final test group. Therefore, none of the test group proper had enhanced their performance due to familiarity with the test method. The test group involved in this study were tested almost immediately upon arrival at the school (see 5.3.6) and so they had not had the chance to participate in any previous, similar in-class activities, and therefore gain from the subsequent learning effects.

(iv) Personal factors: A number of testees' personal and psychological factors can decrease reliability, such as poor health, fatigue, lack of interest or motivation, cognitive style and test-wiseness. Other factors such as sex, age and learning background may also have effect but none of these factors were researched to any depth greater than that covered by the test candidate background questionnaire (see Appendix I). The test group was however, too small for factors such as age or sex to have any significant influence on the overall results.

It can therefore be concluded that an adequate degree of reliability and validity was achieved as the test was designed and implemented with due attention to aspects (i) - (iv) above.

5.3.6 Test Group Profile

A 'test candidate background questionnaire' (see Appendix I) was distributed to 30 native Japanese students enrolled on a 3-month preparation course for the U.C.L.E.S. First Certificate of English (FCE) examination. The course was held at Scanbrit School of English in Bournemouth, England. The questionnaire addressed personal factors that could
potentially influence (or skew) the results of the EA survey. All of the students selected were from the Tokyo area and thus shared the common Japanese dialect of 'kokugo' (see Ashton 2006: 21-22). They arrived on approximately 15th - 20th August 2005 and were immediately surveyed and tested. They eventually returned to Japan by the end of December after completing their examinations. From the 30 students who completed the questionnaire, 20 were selected for testing (13 female, 7 male). From their responses given on the questionnaire these students appeared to be the most homogenous in linguistic background and linguistic/phonetic skills. In general:

• The entire test group had studied English at high school and then at university for approximately one year. They had never studied English abroad before going to England or at evening classes, nor had they used it at work.

• None of them held any additional English 'paper qualifications'.

• All of them had travelled to English speaking countries (not England) but for no longer than two weeks.

• None of them had any English-speaking friends or colleagues as such. They seemed to use English mainly to speak to their teachers, host families or other students.

• None of them had ever studied any other languages.

According to the results of the placement tests used to determine eligibility for the FCE course, the group would most likely obtain an average of grade 'C' in the final examination. This assumption was made before the course had started, prior to any form of instruction at the school. This group of learners would most likely fall within Band 5 (or 'intermediate level', to use
the common terminology of EFL textbooks) on the English-Speaking Union’s nine-level scale of IL proficiency.

5.3.7 Pilot Testing

Parts I and II of the test were first pilot-tested on a group of five Japanese learners, independent of the twenty-member test group but of similar IL development and linguistic backgrounds. This was ascertained through responses given on the same test candidate background questionnaire as detailed in 5.3.6. Observations were made on the effects of the following aspects of the test and the testing process on the pilot test group:

(i) Potentially unknown lexical items (and their explanations): The testees did not seem deterred by any of the potentially unknown lexical items involved in the test as long as their meanings were explained beforehand, and in the case of Part I, there was an appreciation that both answers in the pairs of test words were contextually possible. One should note that the meanings of unknown items were taught without the tester using the items themselves. Pre-teaching of pronunciation, which would of course have aided the testees, potentially skewing the results, was thus avoided.

During Part II testing, the testees did not seem to hesitate, stop to ponder or generally seem less confident in pronouncing less familiar items as long as they did not appear orthographically complex (see 5.3.5 (i)). The pictures

---

8 'Band 5' is defined by the ESU (1993) as "uses the language adequately in familiar situations. Rather frequent problems but usually succeeds in communicating general message."
provided above each test word appeared to clear up most doubts as to their meanings.

It was also found that during Part II, the test sentences produced did not vary much from testee to testee, increasing reliability considerably. For example, in response to the prompt for initial [f] as in 'funny', the sentence "Tim is very funny" was produced by many of the testees. Similarly for the sequence [ʃeɪʃ] as in 'shave', the testees produced "Tim shaves every day". This fortunate phenomenon boosted the reliability of the results, as alternative (and reliable) samples of each test speech sound could be found to cross-check quality of articulation (see 5.4.3).

(ii) Time restrictions on reception and spontaneous production: During Part I, the time lapses between the test sentences on the tape seemed to allow the testees too much time to mentally prepare for each one. After pilot testing they were therefore shortened to almost half the time (approximately five seconds).

During Part II, it was found advantageous to place pressure on the testees and 'hurry them along' through each item as this increased spontaneity and the natural flow and rhythm of their sentences. During both parts of the test, shortening time lapses between test items seemed logical as the testees would perhaps be under similar pressure in a more realistic communicative situation.

It was also observed that, during Part II production, if allowed to repeat their responses, articulation noticeably improved the second time. This appears to illustrate the notion that IL is in a constant state of flux and instability (see 2). Due to the learning effect through repetition of a test item, articulatory ability seemed to improve within seconds! Therefore,
only the first attempt at articulation of each item was accepted for marking. This appeared more representative of the testees' state of IL prior to testing than successive attempts after which learning took place.\(^9\)

(iii) **Overall effects of the test conditions**: Most of the test group balked at the idea of sitting down to do a 'formal test' during Part I testing, and speaking into a microphone during Part II. In order to reduce anxieties as much as possible, effort was made by the examiner to make the testing atmosphere as informal and as light-hearted as possible, without detracting from the seriousness of the test itself.

Upon consideration of these observations, the following procedures were repeated during the succeeding test phases that this study is based upon:

- Explanation of potentially unknown lexical items.
- Restriction of time allowed to respond.
- The opportunity to respond only once.
- 'Deformalising' of the testing environment.

5.4 **Identification of Errors: Design and Implementation of a Scoring System Suitable to This Study**

The second stage of the EA framework involves the establishment of procedures for quantifying or scaling observations of performance, i.e. the

---

\(^9\) This observation seemed to correlate with the initial assumption that during Part I, testees would be more likely to choose the correct answer if each example on the tape were repeated a second time (see 5.3.3). Both reception and production may improve through the learning effect of repetition, which would reduce test reliability. A further investigation into this phenomenon and the effects of varying degrees of pressure on the testees' abilities would prove interesting.
construction of marking protocols, keys and categories to be used by the examiners or markers. An objective scoring system was applied to the aural responses gathered during Part I testing and a subjective scoring system to the oral responses from Part II. The reliability and validity of these systems were also considered.

5.4.1 Part I: Objective Scoring System

During this part of the test, the candidate either recognised the test sound (as heard on the tape) or not and marked the answer sheet accordingly. The answer sheets were then collected and the total numbers of correct and incorrect responses for each item added up into a score. As the responses could only be marked 'correct' or 'incorrect', the same results were produced no matter how many markers were employed, provided that all the markers used a uniform answer key (see Appendix II for a copy of the Part I marker's answer key used during this study). This system was therefore 100% reliable, also displaying high inter-marker reliability (Underhill 1987: 8, Lado 1961: 28-29, Harrison 1983: 110, Hughes 1989: 19).

5.4.2 Part II: Subjective Marking System

Part II involved judgement, on the part of the examiner, of the standard of each testee's production of the speech sounds under test. Harrison (1983: 22, 111) points out that to maintain validity, it is necessary to devise a subjective marking system in the early stages of test development as it is, in principle, a forecast of the testees' performance based on their estimated
stage of IL development. For easy final comparison with the 'correct' or 'incorrect' responses of Part I, a nominal scale was formulated involving a dichotomy of two defined grades: 'Satisfactory' and 'Poor' (to correspond with 'correct' or 'incorrect' respectively). The definitions of these two grades were influenced by Kenworthy's (1987: 13) interpretation of the concept of 'Intelligibility':

'Satisfactory': Intelligible. The quality of articulation of the test speech sound is such that the meaning of the word that it is part of is clear and unobscured (disregarding contextual errors of grammar, syntax, etc.)

'Poor': Unintelligible. The articulation of the sound renders the word incomprehensible, or comprehensible only from guesswork based on its context.

The definitions of these two grades reflect the supposed level of IL, and therefore the quality of speech sound production expected from the testees profiled in 5.3.6. Note that there is no reference to 'native speaker-like' production, inappropriate as the testees would presumably be incapable of such precision at this stage of IL development. The 'satisfactory' grade thus establishes a suitable 'cut-off' so that non-native-like production may still be regarded as successful or adequate (Davies 1990: 18). The above two-point scale also makes the task of grading the results of Part II by ear, easier. A more complex scale would be more difficult for a marker to apply without phonetic training or the use of equipment such as a spectrograph or specialised computer software.

Subjective scoring does not result in high reliability as different markers

---

10 'Intelligibility' presupposes the participants, i.e. it has as much to do with the listener as the speaker (Kenworthy 1987: 14).
can give different scores to the same data. The same marker can also give
the same results different scores if marked on two different occasions.
There is also the question of the background of the marker; familiarity
with the candidates' production can affect the marker's perceptions of
intelligibility (and consequently the above grades of 'satisfactory' and
'poor'). What could be interpreted as better production could merely be
better listening or tolerance ('sympathetic response') on the part of the
marker. For the same reasons the testees' teacher would make an unreliable
marker.

With these problems in mind, the results of this study were first marked by
this researcher (familiar with Japanese learners' articulations in English)
and secondly by an ex-acquaintance\(^\text{11}\) (of this researcher) with little
teaching experience and familiarity with Japanese learners or Japanese
linguistic systems. Reliability could also have been increased with the use of
more than one assessor, but as an added precaution the definitions of
'satisfactory' and 'poor' were discussed prior to marking to ensure uniform
perceptions of the testees' abilities. See Appendix IV for a copy of a Part II

5.4.3 Scoring the Results of the Pilot Test

The responses of the pilot test (see 5.3.7) were scored using the objective
(Part I) and subjective (Part II) systems as detailed in 5.4.1 and 5.4.2. The
following observations were made:

\(^{11}\) The same acquaintance was also to second-mark the results of Part I testing.
(a) No problems or inter-marker discrepancies were experienced during implementation of the Part I objective scoring system due to its uniform nature throughout.

During implementation of the Part II subjective scoring system however, two problems arose:

(b) The recordings of the testees’ production had to be replayed several times to ensure fair and accurate grading of each test speech sound.

(c) Familiarity with Japanese learners' articulation of English sounds generally led to more lenient grading on the part of this researcher (the first marker). The more 'neutral' second marker however found many articulations unintelligible and graded many of them 'poor' in contrast with the first marker's grades of 'satisfactory'. The main areas of inter-marker discrepancy lay in test sounds in their medial and final positions. In these positions, the reduction in articulatory energy involved in the production of these sounds tended to make errors more difficult to identify than when they occurred in more heavily stressed syllables, for example in initial position. (See 6.1.1 (i) for more detailed examples of this phenomenon.) For example:

<table>
<thead>
<tr>
<th>Test Sound</th>
<th>Total Errors (First Marker)</th>
<th>Total Errors (Second Marker)</th>
<th>Discrepancy</th>
<th>Final Negotiated Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>medial [v]</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>final [v]</td>
<td>10</td>
<td>17</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>medial [ʌ]</td>
<td>4</td>
<td>9</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>medial [θ]</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>final [ð]</td>
<td>13</td>
<td>16</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>
In the above (more striking) cases, the total numbers of errors counted by both markers varied from between one to seven. Evidently, the first marker was more 'sympathetic' to the testees' errors than the second marker (see 5.4.2). Other inter-marker discrepancies varied only by one or two errors in total. Such cases were easily cleared up through further inter-marker discussion resulting in the 'final negotiated totals' as above and in fig. 2. Fortunately, due to the highly uniform nature of the testees' oral responses for Part II (see 5.3.7 (i)) many other valid and reliable speech samples were found that could be used to cross-check the articulation of many of the test sounds. For example, amongst the pre-selected test words alone, samples of the following sounds were found:

Initial [ f ] as in 'funny' but also in 'fire' (the test word for final [ ø ]).
Final [ v ] as in 'curve' but also in 'shave' (the test word for the sequence [ ŋei ], etc.

No significant variations were found upon cross-checking of the articulatory quality of these samples, however.

In light of these observations, it was concluded that during successive testing, Part I scoring (as described in 5.4.1) could be easily implemented but that during Part II scoring, the two markers involved should resolve discrepancies (such as the above) through discussion until uniform grades can be decided upon. Also during Part II marking, at least one other sample of as many of the test sounds and sequences should be found within the speech of each testee to cross-check articulatory quality.
5.5 Statement of Findings

For this third stage of the EA framework, the total numbers of 'correct/incorrect' and 'satisfactory/poor' responses from both Parts I and II of the test respectively were cumulated to produce total scores with regard to:

- Individual testees’ erroneous responses to each speech sound and sound sequence tested (see fig. 1 and 2).
- Total errors of each test sound and sequence of the test group as a whole, reproduced from the overall scores in fig. 1 and 2 (see fig. 3 and 4).
- Distribution of sound substitution errors (see fig. 5-8).

5.5.1 Comparison of the Results of Aural (Part I) and Oral (Part II) Testing

Some striking discrepancies were noted in the total numbers of errors made by the individual testees and the total errors concerning each speech sound and sequence:

(i) Total errors made by the individual testees: The error totals for each testee in fig. 1 and 2 reveal that the test group generally produced between 2-6 aural errors and 2-10 oral errors. These totals shed some light on the comparative state of development of their aural and oral phonetic skills and the group’s overall state of IL (see 6.5).
Test Speech Sounds and Sound Sequences

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>5</td>
</tr>
<tr>
<td>2.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>2</td>
</tr>
<tr>
<td>3.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>3</td>
</tr>
<tr>
<td>4.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>5</td>
</tr>
<tr>
<td>5.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>4</td>
</tr>
<tr>
<td>6.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>6</td>
</tr>
<tr>
<td>7.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>5</td>
</tr>
<tr>
<td>8.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>4</td>
</tr>
<tr>
<td>9.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>5</td>
</tr>
<tr>
<td>10.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>3</td>
</tr>
<tr>
<td>11.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>5</td>
</tr>
<tr>
<td>12.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>6</td>
</tr>
<tr>
<td>13.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>6</td>
</tr>
<tr>
<td>14.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>4</td>
</tr>
<tr>
<td>15.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>6</td>
</tr>
<tr>
<td>16.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>5</td>
</tr>
<tr>
<td>17.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>6</td>
</tr>
<tr>
<td>18.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>5</td>
</tr>
<tr>
<td>19.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>2</td>
</tr>
<tr>
<td>20.</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>11</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>20</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total Errors for Each Speech Sound and Sound Sequence (out of 20 Responses)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall total errors (out of 400 responses)

Key: ✔ = Correct responses
     ✗ = Incorrect responses

Fig. 1: Individual testees’ responses and overall totals of errors for each test item elicited during Part I (Aural) testing.
### Test Speech Sounds and Sound Sequences

#### Key:
- **S** = ‘Satisfactory’ responses
- **P** = ‘Poor’ responses

#### Fig. 2: Individual testees’ responses and overall totals of errors for each test item elicited during Part II (Oral) testing.
Fig. 3: Total correct/incorrect responses with regard to each test item elicited from the whole test group during Part I (aural) testing.

Fig. 4: Total satisfactory/poor responses with regard to each test item elicited from the whole test group during Part II (oral) testing.
Distribution of Sound Substitution Errors in the Cases of Initial, Medial and Final [\textipa{\textalpha}] and Final [\textipa{v}] (%)

**Fig. 5**: Initial [\textipa{\textalpha}] sound substitution errors (the voiced plosive [d] was used in preference to the voiced fricative [z]).

- [z] 33%
- [d] 67%

**Fig. 6**: Medial [\textipa{\textalpha}] sound substitution errors (the voiced fricative [z] was used in preference to the voiced plosive [d]).

- [d] 40%
- [z] 60%

**Fig. 7**: Final [\textipa{\textalpha}] sound substitution errors (the voiceless correlate [\textipa{\texteta}] was used in preference to the voiced plosive [d]).

- [d] 13%
- [\textipa{\texteta}] 87%

**Fig. 8**: Final [\textipa{v}] sound substitution errors (the voiceless correlate [f] was used in preference to the voiced stop [b]).

- [b] 7%
- [f] 93%
(ii) Aural errors: Listed below are some of the more significant variations in the total numbers of aural errors observed concerning the test sounds:

(a) No errors were made in the perception of initial [f] but eleven with final [f].
(b) Five errors were made with initial [v], seven with medial [v], but only three with final [v].
(c) None of the testees discriminated medial [r] correctly but only two errors were observed with initial [r] and four with final [ə] (‘orthographic final ’r’”).
(d) Four errors were observed with initial [θ], five with medial [θ] but eleven for final [θ].
(e) No errors were made at all in the perception of the sequences [wi], [wi], [wɛi], [ʃɛi], [ʃɛ] and [ti].

(iii) Oral errors: Similar discrepancies were found in the testees’ oral results:

(a) Only one error was made with initial [f] but five with final [f].
(b) Five errors were observed with initial [v] and two with medial [v], in contrast with fifteen for final [v].
(c) Initial [r] totalled six errors, medial [r] totalled five, but final [ə] totalled thirteen.
(d) Six errors were observed with initial [θ], five with medial [θ] but fifteen for final [θ].

The differences in numbers or oral errors observed with regard to the sound sequences tested, were not as striking as those mentioned in (i).

(iv) Sound substitution errors: Striking variations were also found concerning the oral ’sound substitution errors’ observed in the cases of
initial, medial and final [ʊ] and final [v].

(a) Initial and medial [ʊ]: the voiced plosive [d] was preferred to the voiced fricative [z], but in medial position the opposite was observed.

(b) Final [ʊ, v]: in the majority of final [ʊ] errors, the voiceless correlate [θ] was used. The same was noted with regard to final [v] errors, in which voiceless [f] was used in preference.

Upon analysis of the data in fig. 1-8, the following observations were made:

5.5.2 The Consonants [f, v, θ, ʊ] and Final [ə]

(a) Initial and medial [v, θ, ʊ]: More errors occurred in reception when syllable medial than initial; none of the testees perceived [θ] correctly in medial position. More errors occurred in production when syllable initial than medial.

(b) Final [f, v, ʊ]: In production, the most errors occurred in syllable final position. Such was the case in reception for final [f, ʊ] but the least errors were made with final [v], i.e. less than initial or medial [v].

(c) [v, ʊ] Sound substitution errors: In production, substitutions of sounds resembling [d] or [z] tended to occur in the case of poor articulation of initial and medial [ʊ], but in final position misarticulations resembled [θ]. It was also found that final [v] was very often substituted with a sound minimally different in manner of articulation from [f] (see fig. 5-8).

(d) Final [ə]: A large number of errors were observed in production but not as many in perception of final [ə] (orthographically represented as 'r').
5.5.3 Consonant - Vowel Sequences

(a) [ w ] + vowel: In production, [ w ] + vowel sequences were found to be (from left-to-right) most-least problematic in the following order of difficulty: [ wο ], [ wi ], [ wοє ], [ wi ], [ wєι ]. In reception however, errors were only made in the cases of [ wο ] and [ wοє ], and in the same number as each other. [ wi ], [ wi ] and [ wєι ] were all perceived 100% correctly by the whole group.

(b) [ є ] + vowel: In production, the sequence [ єє ] proved slightly more troublesome than [ єєι ]. Both were perceived 100% correctly by the whole group however.

(c) [ тʰ ] + vowel: There was no difference in number of satisfactory articulations between the sequences [ тʰі ] and [ тʰuw ]. [ тʰuw ] however, proved somewhat problematic in perception but no errors were made in the perception of [ тʰі ].

5.5.4 Relation of EA Results to CA Hypotheses

From the CA hypotheses stated in Ashton 2006: 39-40, the following were supported by the results of the EA. They may therefore be explained (albeit speculatively) by means of interlingual psychological notions (see 2):

(i) [ f, v, θ ]: the most problems are experienced in syllable final position.\(^{12}\)

(ii) Final [ ø ]: can be highly problematic in production.

\(^{12}\) Note that these observations arose through intralingual comparison of these English sounds in initial, medial and final positions rather than through interlingual comparison with the corresponding Japanese sounds.
(iii) \([ t^h] + \text{vowel}\): negligible or no difference in difficulty is experienced between sequences of \([ t^h] + \text{vowel}\) (at least in production).

Some of the EA results proved contrary to, or more conclusive than, the CA hypotheses. These may still be explained by different interlingual means however, or by examination of the acoustic properties of some of the sounds.

(iv) \([ v, \partial, \mathfrak{t}]\): varying degrees of difficulty are experienced in both reception and production when syllables initial and medial.

(v) \([ w ] + \text{vowel}\): \([ w\partial ]\) is by far the most problematic sequence from those tested, both in reception and production. A hierarchy of difficulty in production was also established for all five sequences tested (from the EA data) but with the sequence \([ w\varepsilon ]\) as the least, not the most (as predicted by the CA) problematic.

(vi) \([ \mathfrak{s} ] + \text{vowel}\): the sequence \([ \mathfrak{s}\varepsilon ]\) is slightly more troublesome in production than \([ \mathfrak{s}\varepsilon i ]\).

(vii) \([ t^h ] + \text{vowel}\): errors occur in perception of the sequence \([ t^h\varepsilon u]w\) but not in the case of \([ t^h i ]\).

Other observations were made that were not addressed at all by the CA. Other psychological notions (e.g. intralingual) may be applied to this data:

(viii) Reception vs. Production: For all consonant-vowel sequences tested, aural skills are generally more developed than oral. In fact, most of these sequences are not at all problematic in reception, at least at the stage of IL of the test group. Reception of final \([ \varepsilon ]\) was also more advanced than production.

(ix) Sound substitution errors: Substitutions of sounds resembling \([ z ]\) or \([ d ]\) occur when \([ \partial ]\) is syllable initial or medial but do not tend to occur
when syllable final. In final position, the sound produced (when misarticulated) resembles [θ]. Similarly, final [v] is replaced more often with [f] than the CA-predicted [b].

Section 6 forms the final stage of the EA framework and theoretically analyses the data generated by the EA survey *a posteriori*, in terms of the interlingual and intralingual psychological theories of language learning and communication as described in 2 and 3 and the acoustic properties of some of the test sounds. An assessment of the IL of the test group is also made (based on the data displayed in fig. 1 and 2) as well as general conclusions relating to the present study as a whole.

### 6. Theoretical Analyses and Applications of the EA Survey Data and General Conclusions

In this chapter, the following points are discussed:

(i) The empirical data generated by the EA survey is placed in a theoretical perspective in terms of the interlingual and intralingual/developmental psychological theories of language learning and communication related in 2 and 3 respectively. Where appropriate, the acoustic properties of some of the test sounds are considered.

(ii) The aims of this study are addressed and proposals are made based on the empirical EA data. These may be added as supplementary information to Nilsen and Nilsen's (1971) and in particular Kenworthy’s (1987) descriptions of the English L2 problems that native Japanese learners experience in their already published works (see Ashton 2006: 6-10). The implications of this information on classroom teaching are also discussed.

(44)
(iii) The state of IL development of the test group is assessed, based on the individual and overall learner data tabulated in fig. 1 and 2. The notion of fossilisation is also considered (see 4.)

(iv) General conclusions are drawn regarding the methods employed during this study and areas of further research. The relationship of CA and EA as studies of IL and their underlying psychological notions are also discussed, with reflection on the present study.

6.1 Theoretical Analysis of the EA-Generated Data

Interlingual and intralingual psychological theories, as well as factors such as the acoustic properties of some of the test sounds, may be related to the EA survey data (see 5.5). One should note that the following comments are merely statements of what the data appears to suggest.

6.1.1 The Consonants [ f, v, η, ɔ ] and Final [ ø ] (See 5.5.1)

(i) Initial and medial [ v, η, ɔ ]: Dissosway-Huff et al (1982) predicted, from the descriptive data of their English-Japanese CA, that the English consonants [ l, η/əɹ ] would be more difficult to acquire in syllable final position than initial or medial,13 due to relatively fewer post-vocalic

---

13 One should note that in Standard Southern Spoken British English, [ η ] does not occur in final position but that final orthographic 'r' often corresponds to [ ø ]. In American English however, (Dissosway-Huff et al conducted their research in Indiana, USA, involving Native American English speakers), the sequence [ əɹ ] occurs. Thus British English 'fire' [ faɪə ] is generally pronounced [ faɪəɹ ] in American English.
consonants in Japanese (see Ashton 2006: 37). The closest approximating sound (in articulatory terms), Japanese [τ], also does not occur in final position. In their research however, also that of Mochizuki (1981) and Sheldon and Strange (1982), it was found that Japanese learners were able to distinguish [l, ɹ] with greater accuracy in final position\(^{14}\) than initial or medial. An examination of the acoustic properties of these sounds in the speech of native (American) English speakers revealed that their combined mean duration varied when syllable initial, medial and final (excluding consonant clusters). These were:

- Syllable initial: 108 ms
- Syllable medial: 121 ms
- Syllable final [l, ɹ]: 195 ms

A further investigation of the same phenomena in the speech of native British English speakers would be necessary before such figures may be accurately related to the findings of this study. The above figures will therefore only be regarded as a very general guideline. However, these findings tentatively offer an acoustic explanation as to why Japanese learners have a greater ability to perceive final [l, ɹ] more accurately than initial or medial [l, ɹ]. Similarly, medial position [l, ɹ] should be easier to perceive than initial. (One should also note that preceding vowel length or quality vary with their following consonants. Thus, in the case of medial position consonants, the preceding vowel is also a clue to discrimination).

\(^{14}\) The authors make no comment as to whether this refers to genuinely absolute final positions or perhaps on whether their results reflect the concentration of teaching; 'difficult' points are often taught more.
Henly and Sheldon (1986: 507, 516) label the above acoustic phenomenon 'the Duration Hypothesis'. On the principle that [ l, ɾ] do not exist in Japanese, the hypothesis could theoretically also be extended to other such sounds, e.g. [ f, v, ð ]. These would appear reasonable enough assumptions.

In the English-Japanese CA conducted during the previous study (see Ashton 2006: 37-38), it was noted that final [ f, v, ð ] were longer in duration than when initial or medial (no comment was made on differences between initial or medial position however). Unfortunately, the EA data from the present study concerning initial and medial [ v, ɾ, ð ] contradicts the Duration Hypothesis. In each case, more errors were observed in the perception of medial [ v, ɾ, ð ] than initial. An alternative explanation must therefore be rendered.

In both Part I (aural) and Part II (oral) of the EA test, the test words all followed very similar stress patterns (see below):

<table>
<thead>
<tr>
<th>Initial Position</th>
<th>Medial Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ v ]:</td>
<td>'vowels (aural and oral)</td>
</tr>
<tr>
<td>[ ð ]:</td>
<td>'those (aural)</td>
</tr>
<tr>
<td></td>
<td>'then (oral)</td>
</tr>
<tr>
<td>[ ɾ ]:</td>
<td>'right (aural and oral)</td>
</tr>
<tr>
<td>[ v ]:</td>
<td>'rover (aural and oral)</td>
</tr>
<tr>
<td>[ ð ]:</td>
<td>'breathing (aural)</td>
</tr>
<tr>
<td></td>
<td>'heather (oral)</td>
</tr>
<tr>
<td>[ ɾ ]:</td>
<td>'pirate (aural and oral)</td>
</tr>
</tbody>
</table>

'= position of main/primary stressed syllable. The other syllables are lightly stressed by comparison.

As can be seen from the above, [ v, ɾ, ð ] occur in heavily stressed syllables in initial position but lightly stressed syllables in medial position. The greater articulatory energy (stressing) involved in the production of these
sounds (by the native speaker on the tape recording used during Part I recording) may have made them easier to distinguish when syllable initial than medial ('medial' here means 'initial in a relatively less stressed syllable'). None of the testees perceived medial [Ʌ] correctly, making it a particularly extreme case.\(^{15}\) From another point of view, the Japanese syllable-timed sound system may have interfered more in the perception of syllable medial sounds, in which case the errors could be interlingual in origin.

In oral testing (Part II) more errors seemed to occur in initial than medial position. One must not rule out the (albeit unlikely) possibility of light-stressing of syllable medial speech sounds having a similar effect on the marker's perception of syllable medial errors. Most of the testees correctly reproduced the stress patterns of the test words, i.e. heavy stress on syllable initial in all cases (as above). Possibly the errors may merely have been more noticeable to the marker due to their occurrence in (the more heavily stressed) syllable initial than medial. On the premise that 'syllable medial' tends to mean 'initial in the next syllable' it is conceivable that the degree of error of \([\upsilon, \lambda, \delta]\) in both initial and medial positions was, in fact, the same or very similar.

It could therefore be concluded that correct production of English syllable stress and stress rhythm can unfortunately highlight, but fortunately obscure (facilitating communication) Japanese learners' errors in pronunciation, at least as far as the listener is concerned. By the same token, it can

---

\(^{15}\) A true test of the validity of the Duration Hypothesis would (amongst other things) have to involve different test words in which the test speech sounds all occurred in heavily stressed syllables in both initial and medial positions.
also hinder or facilitate learners’ perception of English sounds, and in turn communication skills.

**(ii) Final [ f, v, .signup ] and final [ œ ]:** Final [ v ] appears to be the only test sound that has produced data corresponding to the Duration Hypothesis (see (i)). The longer duration of the acoustic signal could possibly account for the testees’ greater ease in discriminating final [ v ] from initial or medial.

Final [ .signup ] could be more difficult to distinguish due to the reduction of energy involved in articulation (reduced when compared to initial, or even lightly stressed medial position) which is interpreted by the learner as an apparent reduction in voicing (see Ashton 2006: 32,37-38). [ f ] is of course voiceless, but nonetheless experiences a reduction in articulatory energy when syllable final; for example, final [ f ] as in ‘thief’ involves less energy than initial [ f ] as in ‘funny’ or medial [ f ] as in ‘careful’. Similarly, final [ œ ] is greatly reduced ([ œ ] can only generally occur in a lightly stressed syllable but, for example, is found more heavily stressed in the diphthong [ œə ] and therefore not easy to perceive, but not difficult to distinguish from final [ l ] as was the case during Part I testing.

In production, considerably more errors were observed with final [ f, v, .signup ] than initial or medial. Japanese learners may exaggerate the tendency for absolute final position fricatives (such as the above) to be followed by a short, parasitic vowel (see Ashton 2006: 38). The testees generally seemed to replace final [ f ] (which appeared in absolute final position in their sample sentences) with the Japanese sequence [ Φu ]. Final [ œ ] was substituted in most cases with a sound resembling Japanese [ a ] and was also articulated as a separate syllable (see Ashton 2006: 38). Both of these phenomena were
predicted by the CA and so may be labelled interlingual errors.

(iii) [v, ɔ]: Sound substitution errors: The absence of [ɔ] in Japanese and the descriptive data generated by the English-Japanese CA suggest that errors in the production of [ɔ] could be caused by interference from Japanese [d]. This appeared to be the case in the majority of errors in syllable initial (see fig. 5), a large number in syllable medial (see fig. 6) and a few in syllable final (see fig. 7). Other errors in these positions resembled [z]. [b] was observed in all initial and medial [v] errors but only a small amount in final position (see fig. 8). As [b, d, z] all occur in Japanese, the source of these errors could be interlingual.

However, the majority of final [v, ɔ] errors resembled [f] and [θ] in the cases of [v] and [ɔ] respectively. These occurrences could possibly have been the results of exaggeration of the reduction in energy involved in final position articulation, which accounts for a reduction in voicing; [f] and [θ] are the voiceless correlates of [v] and [ɔ] respectively, in articulatory terms. As articulatory energy is lower at the end of the syllable, the sound is interpreted as voiceless [θ]. The learner actually produces fortisness as compensation for the voicing in [ɔ] (lenisness). One must also not rule out the possibility of 'spelling pronunciation errors' (see 3) in the case of [ɔ]; both [θ] and [ɔ] are orthographically represented by 'th' in English, which many Japanese learners overgeneralise as [θ] in pronunciation.

In either case, these errors could be the result of inadequate learning of the relevant L2 articulatory rules or false L2 hypothesising as to the pronunciation of [v, ɔ] in which case the errors would be intralingual.
6.1.2 Consonant-Vowel Sequences (See 5.5.2, Ashton 2006: 39-40)

(i) [ w ] + vowel: Interference from differences in syllabic structure appeared to be less of a problem than articulatory differences between these English sequences and their Japanese counterparts. These observations did not so much contradict the predictions of the CA as redefine or redirect them. For example, the EA concluded that the sequence [ wʊ ] was the most problematic of the five sequences examined whereas the CA was inconclusive on that point.

The errors nonetheless appear to be interlingual; e.g. [ wʊ ] was clearly articulated as Japanese [ uɪ ] and the sequence [ wi ], a single syllable, was articulated as Japanese [ ʊi ], which spans two syllables. Interestingly, fewer errors were observed in cases in which greater degrees of syllabic interference, caused by greater differences in English-Japanese syllabic structure, were expected. For example, the least errors were observed in the production of [ wɛi ] (one syllable) which can be contrasted with Japanese [ ʊei ] (potentially three syllables). Either the test group on the whole had a greater mastery of English syllable stress and sentence rhythm than initially expected or the difference in syllable structure between the two languages are, in these cases, so striking to the learner that more care is taken not to allow errors to occur.

(ii) [ f ] + vowel: As with the sequences of [ w ] + vowel, the predictions of the CA were more redirected than contradicted. The sequence [ ʃɛ ] proved more problematic on production than [ ʃɛi ], in which syllabic interference was predicted. This showed that syllabic structure was not as much a cause of interference as articulatory differences. In these cases, errors with [ ʃɛ ]
and [ jsei ] resembled Japanese [ se ] and [ sei ] respectively and so it can be concluded that the source of errors was interlingual.

(iii) [ tʰ ] + vowel: No difference in the number of errors made when producing the sequences [ tʰuw ] and [ tʰi ] were observed in the EA data. The sources of error in each sequence possibly differ from those predicted by the CA, however. It was observed that errors in the case of [ tʰuw ] resembled the Japanese sequence [ tsuᵢ ], therefore these errors may be interlingual in origin. With regard to [ tʰi ], errors seemed to resemble a sequence [ tsi ], contrary to the CA-predicted [ ūi ] . [ tsi ] does not occur in Japanese but is nonetheless quite conceivable; both the consonant [ ts ] and the vowel [ i ] exist in the Japanese sound system. [ ts ] rarely occurs in initial position in modern English (e.g. 'tsetse fly') but is common post-vocally (e.g. 'lets', 'cats', etc.). Learners are unlikely to be aware of the occurrence and rarity, of English initial [ ts ] at this stage of IL and so the source of error is more likely to be interlingual (further interference of Japanese [ ts ] , as in the case of [ tʰuw ] ) than intralingual (false hypothesising of the pronunciation of [ tʰi ]).

Interference of Japanese [ ts ] may also explain the difference in number of errors that were made in the perception of [ tʰi ] (0%) and [ tʰuw ] (35%). In Part I of the test, [ tʰuw ] was contrasted with the sequence [ suw ] . Japanese [ ts ] resembles English [ s ] and so could cause confusion between the test sounds, but not necessarily in the discrimination of [ tʰi ] and its contrasting sequence [ ŋi ] , as used in the test.
6.1.3 Summary: Interlingual vs. Intralingual/Developmental Errors

It would appear from the above analysis that most of the errors observed during the EA survey may have been interlingual in origin, caused either by interference from 'single' Japanese speech sounds or from the Japanese syllabary system. However, Japanese syllabic interference appeared to cause differences in the number of errors concerning consonants in varying syllabic positions (in particular, perception in initial and medial positions; although it should be noted that consonant clusters were not investigated) and on the whole, less errors in the various sequences of consonants and vowels than expected. Very few of the errors seemed likely to have intralingual/developmental sources (only some of the errors observed in the production of final [v, ɔ]) or differ in number due to variations in acoustic properties.

6.2 Aims of the Present Study

The data collected throughout the present study and its analysis in 6.1 allows the aims of the previous CA study described in Ashton 2006 to be addressed. To recap, these aims were to find out:

i. Whether certain English consonant sounds can be more or less troublesome to Japanese learners in their initial, medial or final positions (see 6.2.1).

ii. Which sound substitutions occur more or less often and in which syllable positions in the L2 English production of Japanese learners (see 6.2.2).
iii. Whether sequences of a given consonant and a selection of vowel sounds can be more or less problematic to learners than each other (6.2.3).

iv. To provide supplementary comments that could be added to references to the above points already made in Nilsen and Nilsen's (1971) and Kenworthy's (1987) published texts (see 6.3).

6.2.1 English Consonants in Initial, Medial and Final Positions

(i) Oral skills: In production, it is debatable whether there is any difference at all in degree of articulatory difficulty between the consonants tested in syllable initial and medial positions as, after all, 'medial position' often means 'initial in the next syllable' (see 6.1.1 (i)). If the learner can develop a mastery of English syllable stress and stress-timed rhythm however, errors can at least be made less noticeable to the listener, thus facilitating communication.

Final position does, however, appear considerably more problematic apparently due to the learner's exaggeration of the tendency for final fricatives to be followed by a short parasitic vowel; possibly a result of interference from the Japanese syllabary system, in which very few post-vocalic consonants occur.

(ii) Aural skills: In perception, the data and their possible explanations are more varied. Interference from the Japanese syllable-timed system could hinder the perception of English sounds in lightly-stressed syllables, e.g. syllable medial in the case of the sounds tested. Japanese syllable-timing could also cause interference in production and reception of the reduced vowel [ə].
Final [f, s] appeared the hardest to discriminate, possibly due to the reduction in articulatory energy involved (accounting for an apparent reduction in voicing in the case of final [s]). Final [v] however, could be easier to perceive due to its longer acoustic signal than when initial or medial, according to the Duration Hypothesis\(^{16}\) (see 6.1.1 (ii)).

### 6.2.2 Sound Substitution Errors

(i) **Initial, medial and final [z]:** In initial position, errors tend to resemble [d] and in medial position [z]. Both of these cases of errors appear to be of interlingual origins. Errors in final position seem to result in a sound close to English [θ], which could have intralingual sources (see 6.1.2 (iii)).

(ii) **Initial, medial and final [v]:** Initial and medial [v] were in all cases of error replaced with the sound [b] (possibly from interlingual causes) but a sound approximative to [f] in the majority of cases in final position. As with final [s], these errors could have intralingual sources (see 6.1.2 (iii)). Further research could uncover other English sounds that suffer different sound substitutions in different syllable positions when produced in error.

### 6.2.3 More or Less Problematic Consonant-Vowel Sequences

More errors appear likely to occur in sequences in which Japanese 'articulatory' differences are expected to interfere than syllabic differences.

\(^{16}\) Further research is needed to determine whether the Duration Hypothesis could apply to the perception of other English consonants in initial, medial and final positions.
This assumption would appear applicable to both aural and oral skills. L1 interference seems to be a major source of error in this area. Further CA/EA research would no doubt expose other English consonant-vowel sequences that were more problematic than others to Japanese learners for similar reasons (see 6.1.2 and 6.1.3). Hierarchies of difficulty in sequences of a given consonant and various vowels, as established by the data from the present study, would appear to be as follows (left-right = most-least problematic; see also 5.5.2):

- \[w\] + vowel: [\(\text{wu}\)], [\(\text{wi}\)], [\(\text{wu}\varepsilon\)], [\(\text{wi}\)], [\(\text{we}\varepsilon\)]
- \[\text{ʃ}\] + vowel: [\(\text{s}\varepsilon\)], [\(\text{s}\varepsilon\i\)]
- \[\text{t}^h\] + vowel: [\(\text{t}^h\text{uw}\)], [\(\text{t}^h\text{i}\)] (in perception only)

### 6.3 Supplementary Information to Existing Published Works

Below are extracts from Kenworthy’s (1987) publication “Teaching English Pronunciation” as discussed in Ashton 2006: 6-12. The observations made in 6.2 may be added as supplementary comments to references already made by Kenworthy to certain aural and oral problems experienced by Japanese learners of English. The same comments may also be added to Nilsen and Nilsen’s (1971) text but Kenworthy’s publication is referred to here as it is, in general, more detailed than Nilsen and Nilsen’s and provides a context within which the comments generated by the present study may be applied more easily. The extracts below are from Kenworthy (1987) pages 149-152. In each case, Kenworthy’s reference appears first in standard print. Additional comments (provided by this researcher) follow in italics:
Problems with consonants: The following consonants do not occur in Japanese: \[ v \] as in ‘van’, \[ f \] as in ‘fan’, \[ θ \] as in ‘soothe’, \[ η \] as in ‘run’.

It is unlikely that the above consonants will prove more or less problematic for the learner when syllable initial or medial.

Communication strategy: Mastery of English syllable stress and sentence rhythm may enable the learner to obscure misarticulations from the listener when the above sounds occur in lightly-stressed syllables, thus enhancing communicational clarity. For example, medial \[ v \], as in ‘rover’, will seem less like an error to the listener as it occurs in a lightly-stressed syllable, than ‘revenge’ where it occurs in a heavily-stressed syllable. By the same token however, the Japanese learner may have more problems in discriminating the above sounds when they occur in lightly-stressed than heavily-stressed syllables. One should also note that the Japanese sound system is syllable-timed\(^{17}\), in contrast to the English stress-timed system. This may cause interference in the above areas.

In final position however, learners may add extra final vowels in production as post-vocalic consonants are rare in Japanese. This may result in an exaggeration of the parasitic vowels that occur when the sounds are absolute final position in English. There is also a reduction in articulatory energy involved in final position, which accounts for an apparent reduction in voicing of some sounds (e.g. \[ v, θ \]) and which may also cause confusion in discrimination. Final \[ v \] could be an exception however; its longer acoustic signal may in fact make it easier to perceive.

'Sound Substitution' Errors: The sounds \[ z \] or \[ d \] will be used for \[ θ \].

---

\(^{17}\) Kenworthy gives further explanation of the Japanese syllable-timed system in a later section.
[θ] tends to be substituted with [d] in initial position and [z] when medial. Both sounds could also occur in error in final position. [θ] can also occur in errors of final [θ] as an exaggeration of the reduction in articulatory energy involved (that accounts for an apparent reduction in voicing; see above), or simply overgeneralisation of the pronunciation of orthographic 'th' as [θ]. Similarly, final [v] may be replaced with a sound resembling [f].

**Problematic consonant-vowel sequences:** The following consonants occur in Japanese and can be transferred to English with little or no need for modification: [p, b, t, d, k, ɡ, ʃ, s, z, ʃ, m, n, η, w, j, h]. However, some of these consonants only occur before particular vowels in Japanese, so learners may have problems when they try to pronounce an 'unfamiliar' consonant-vowel sequence. Here are some potential trouble spots:

**a)** The only vowel which commonly follows [w] as in 'wet' in Japanese is [a], so whereas 'wax' will not be a problem, 'win', 'wait', 'would', 'we', etc. may be.

*From the above examples, the sequences [wɔ] as in 'would' and [wi] as in 'win' are potentially more problematic than the others. Their closest approximating Japanese sounds differ more in articulatory terms in these two examples than the others, causing greater interference in English production. Interference from the Japanese syllable system may also result in other sequences articulated as more than one syllable.*

**b)** The sequences [tʰi] and [tʰu] do not occur in Japanese. Watch out for problems with words like 'team', 'two', etc.

*Although it is unlikely that either sequence could cause more problems to the learner than the other in production, the Japanese sound [ts] could
interfere more in the perception of $[t\,\text{uw}]$.

c) $[\mathcal{f}]$ does not occur before $[\mathcal{g}]$ as in 'bed' or $[\mathcal{gi}]$ as in 'made'; possible problems will be 'shade', 'shell', 'shame' etc.

Differing degrees of difficulty are unlikely in the cases of the sequences $[\mathcal{e}]$ and $[\mathcal{es}]$. As above however, $[\mathcal{es}]$ is potentially more troublesome due to greater articulatory differences from its closest Japanese counterpart sequence. Interference from the Japanese syllable system may cause $[\mathcal{es}]$ to be pronounced as two syllables, instead of a single syllable but this is generally less of a concern.

From the above, it can generally be concluded that further research could uncover far more detail that could be added to Kenworthy's text. In general, there appears to be a lot of scope for additional comment, especially concerning references to consonantal groups and stress patterns (etc.).

6.4 Implications for the EFL Teacher

The above comments would seem to suggest that greater attention should be paid by the EFL teacher to tuition and practice of English syllable stress and stress rhythm in classes of Japanese learners as this can greatly facilitate the articulation of English speech sounds. Mastery of such aspects of pronunciation can arguably contribute more to making a learner's speech more intelligible. If Japanese syllable-timing is allowed to interfere too much, a word can easily become unrecognisable or incomprehensible to the listener.

The testees involved in this study appeared, on the whole, to have a
reasonable mastery of the English stress system. Failure to adequately reproduce English stress patterns could have resulted in a larger number of observed errors, particularly in the perception and production of the consonant-vowel sequences tested. Such emphasis on the teaching of syllable-stress and sentence rhythm should not be interpreted as an attempt to belittle the practice of other aspects of pronunciation. As can be seen from the results of this study, many problems can arise concerning 'single' speech sounds (and interference from 'single' Japanese sounds). Teachers of Japanese learners should also appreciate that not all errors in English L2 production are the result of L1 Interference. Very often errors may occur due to faulty or incomplete knowledge of how to correctly produce certain sounds, as has also been noted in the course of this study.

6.5 Assessment of the IL of the Test Group

Although the 'level' and linguistic background of the test group as a whole was researched prior to this study (see 5.3.6) it would seem appropriate to reassess the state of development of the group's IL based on the findings of this research. The data generated by the EA survey may be related to the aural and oral phonetic competence of the test group:

6.5.1 Aural Errors

From the data displayed in fig. 1 (see section 5), regarding the testing of aural skills, the group appeared particularly homogeneous. All of the 20 testees produced between 2-6 errors, from a possible total of 20 tested. Some
correlations between certain types of error were observed, that could be regarded as 'typical' to individual testees:

(i) 7 learners made 2 or more errors with medial consonants. This suggests a general weakness in the discrimination of medial consonants, possibly due to the occurrence of the sounds tested in lightly stressed syllables (see 6.1.1 (i)).

(ii) 8 learners made two or more errors with final consonants. This suggests a general weakness in the discrimination of final consonants, possibly due to the reduction in articulatory energy involved (see 6.1.1 (ii)).

(iii) 3 of the 4 learners who made errors with the sequence [wʊ] also did with the sequence [tʰʊw]. This suggests a general problem in discrimination of such high back vowels, possibly due to interference from Japanese [u] (also a high back vowel) in all cases (see 6.1.2 (i)).

6.5.2 Oral Errors

In fig. 2 (oral testing) it can be seen that the errors vary more in number from testee to testee; 2-10 errors per testee from a possible total of 20 tested. Further correlations, suggestive of 'typical' errors, were observed:

(i) 4 learners made 2 or more errors with initial consonants. These errors were perhaps merely more noticeable to the marker due to their occurrence in heavily stressed syllables (see 6.1.1 (i)).

(ii) 13 learners made 2 or more errors with final consonants. This suggests a general weakness in the production of final consonants due to an exaggeration of the tendency for English final fricatives to be followed by a short parasitic vowel (see 6.1.1 (ii)).
(iii) 4 learners made errors with both of the sequences [ wu ] and [ wi ]. This would suggest a general weakness in the production of [ w ] + vowel sequences in which interference from closely correlating Japanese single syllable sequences can occur (see 6.1.2 (i)).

6.5.3 Comparison of Aural and Oral Errors

The group made a total of 87 aural errors and 115 oral errors, from possible totals of 400 tested. This suggests that the group’s aural skills were generally more advanced than their oral skills, as can often be the case with many learners (see 5.3.1). It should be noted however that 4 of the learners made more oral than aural errors and another 4 made the same number of aural and oral errors.

Upon comparison, further correlations between the numbers of aural and oral errors were observed:

(i) 10 of the learners who made final consonant aural errors also made final consonant oral errors. 7 of these learners made both aural and oral errors with final [ s ]. This would suggest that a weakness in the production of final consonants (generally found to be the most problematic positioning of the consonants tested) could also indicate a problem in their discrimination, and vice versa. Final [ s ] appeared to be inherently difficult for the group.

(ii) 5 of the learners who made medial consonant aural errors also made final consonant oral errors. This seems to suggest that an aural weakness with medial consonants could also indicate an oral weakness with final consonants, although this correlation could merely be coincidental.

(iii) 3 of the 4 testees who made aural errors with the sequence [ wu ] also
made oral errors with [wɔ]. This appears to be the only significant aural/oral correlation that suggests the skewing of results, as it applies to such a high number of the errors observed. None of the aural/oral errors for the other consonant-vowel sequences correlated.

### 6.5.4 Fossilisation

As the learners had enrolled on a 3-month intensive course as preparation for the U.C.L.E.S. FCE examination (see 5.3.6), it could be assumed that the testees generally wanted to improve their English L2 skills, at least to a level capable of passing FCE. Fossilisation may therefore not have been a significant variable. In a series of studies from 1984-1995 dealing with the perceptual foreign accent, Flege et al conclude that Japanese speakers of English can master sounds absent from the L2 with time and exposure to the L2, thus overcoming fossilisation to a great extent. However, determination of exactly which sounds can be mastered, as well as what kind of phonetic input is needed for successful learning to occur would require further research (Flege et al 1995: 52).

### 6.5.5 Relation of 'ESU Band 5' to the Test Group's Phonetic Competence

In 5.3.6, ESU band 5, or 'Intermediate level' (the estimated stage of IL development of the group as a whole) was defined as "uses the language adequately in familiar situations. Rather frequent problems but usually succeeds in communicating general message." This definition, with
reflection on the data and discussion generated through the present study, still appears to fit the testees' overall state of IL, in terms of phonetic competence. The number of errors produced by each testee (aural: 2-6, oral: 2-10) would suggest frequent problems in communication, as it is generally possible that erroneous perception or production of a single speech sound or sound sequence (if particular types are reported) can confuse the meaning of a word to the point that communication fails. None of the test group made more than 50% total errors, either aurally or orally, which would also suggest that in general, the individual succeeds in communicating a message.

One should note however, that only Part II (oral) of the test provided the testee with a 'familiar situation' ('describing your teacher'; see 5.3.3) in which to communicate, as in the above ESU definition. Part I of the test was only partially contextualised (thus 'unfamiliar'). Even so, the test group generally produced adequate results, relative to the ESU definition. Further, perhaps rapid progress, exceeding ESU band 5 might also be expected as the learners prepare for their examinations.

6.6 General Conclusions

The following conclusions can be drawn with relation to the present study, concerning:

• The methods employed throughout (see 6.6.1)
• Areas of further research (see 6.6.2)
• The relationship of CA and EA as studies of IL and their underlying psychological theories (see 6.6.3)
6.6.1 Methods of Study

The completion of the present study followed a well-known method but there were limitations which were revealed in the course of the work:

(i) Terminology: The field of phonetic analysis is problematic and in dispute (see Ashton 2006: 2-5). There is, for example, considerable disagreement over the fundamental definition of the terms 'phoneme', 'phonotagm', 'phonological unit', etc. Also some analyses (e.g. Kenworthy 1987, Gimson 1962) suggest an English inventory of 45 phonemes whereas others (e.g. Mulder 1974/1978) suggest 26. Such discrepancies indicate serious problems in theory and its application. Clearly, the term 'phoneme' cannot be taken for granted.

The broad transcription of English speech sounds (in fact, referred to in many EFL writers' use of the term 'phoneme') reflects phonological differences which might be analysed in a variety of ways, for example, in terms of communicative competence. Thus, the present study constitutes a contribution to the study of IL at the level of phonetic communicational ability (a concept introduced by the term 'phoneme') but without commitment to any (necessarily controversial) phonological account of that ability. Such a commitment would complicate comparison with other published research.

In light of the above, confinement of this investigation to the study of the articulatory properties of speech sounds in broad transcription, with its subsequent effects on the terminology used throughout, seemed to be entirely appropriate.

(ii) Test group screening: In addition to investigation into the linguistic
background of the test group by means of a questionnaire (see 5.3.6), the speech sounds under investigation could be pre-tested on the test group by means of the same style of test but containing the test sounds in different contextual examples. This may further identify testees that could potentially skew the results, in which case they may be omitted from the test group altogether during successive testing.

A motivational survey of the group prior to testing may also shed more light upon the issue of fossilisation as an L2 performance variable (see 6.5.4) and its possible effects on the performance of the test group. Significantly, ‘fossilised’ learners may then also be omitted from the investigation.

(iii) Cross-checking of test items: The articulation of the test speech sounds was cross-checked with other samples of the same sounds wherever possible (see 5.4.3). Further opportunities to cross-check, purposefully built into both aural and oral parts of the test, would have strengthened the final data generated.

Other improvements to the test methods used are obvious and too numerous to be listed here, for example, the use of expensive equipment such as specialised computer software during the grading of Part II oral responses as well as the use of phonetically-trained markers (see 5.4.2).

6.6.2 Areas of Further Research

Throughout this study, references have been made to areas that would be of significant interest if investigated further, beyond the confines of the

---

18 The consequences of the learning effect on the testees of taking the same type of test twice would need to be taken into consideration (see 5.3.6 (ii)).
present study:

(i) Other aspects that could have been built into the EA framework (see 5.2). For example, determination of the degree of disturbance of errors observed in terms of their effects on communication, and detailed remedial work for learners' errors.

(ii) Determination of which system of cues (e.g. grammatical, lexical, or a knowledge of L2 sound contrasts) the testees use to discriminate troublesome sound contrasts (see 5.3.2).

(iii) Whether testing of the perception and production of the test sounds in isolation, or contextualised to a greater extent (as in 'free speech') would have produced different results to the 'partial contextualisation' used throughout this study (see 5.3.4 (ii)).

(iv) Whether the use of video tape, rather than audio tape, would have produced different results in both aural and oral testing (see 5.3.5 (ii)).

(v) The learning effect through repetition of test items, also subject to varying degrees of pressure (e.g. time restrictions), on the testees' aural and oral skills (see 5.3.6 (ii)).

6.6.3 The Relationship of CA and EA as Studies of IL

The descriptive (CA) and empirical (EA) data generated by this study would seem to support the 'weak claim of CA'; that CA can account, or provide explanation, for some aspects of learner performance a posteriori (see Ashton 2006: 16-17). As can be seen from the theoretical analysis in 6.2, the majority (but not all) of the errors observed throughout the EA survey can be explained as being interlingual in origin.
One should note that the previously conducted CA did not predict all of the interlingual errors observed but could easily have done so had the analysis been more detailed and subtle; for example, more detailed explanation of potential syllabic interference in the cases of consonants in initial and medial positions. Other errors were predicted but the degree of difficulty was left inconclusive by the CA, such as whether the sequence \([ w \alpha ]\) was in fact potentially the most or least problematic to learners of all the \([ w \alpha ] + \) vowel sequences tested (see Ashton 2006: 39).

The EA survey data provided greater clarity to issues such as the above and helped to redirect the analysis of errors. It also uncovered other errors, unpredicted by the CA, such as the substitution of final \([ v, ɛ ]\) with \([ f, θ ]\) respectively (see 5.5.1). This would seem to support the view that EA empirical testing is an essential research tool in the study of learner errors as the L2 performance of the individual learner must be taken into account (see Ashton 2006: 17).

This study can therefore conclude that CA, via the psychological notions of L1 Transfer and Interference, has considerable predictive and explanatory power concerning Japanese learners’ errors in L2 English aural and oral phonetic skills, with the added refinement of empirical investigation. It should be emphasised however, that further research is necessary to attempt to clarify which psychological aspects of Transfer or Intralingual/Developmental processes may be attributed to which of the errors observed.
6.6.4 Aims of the Present Study

The aims of this study, as outlined in 6.2, have generally been satisfied as a result of the successful application of CA and EA methodologies throughout. The further aims of the modification of Nilsen and Nilsen’s (1971) and in particular, Kenworthy’s (1987) well-known texts (and hence the direction of possible EFL materials), as well as the provision of further practical details for the EFL teacher (see 6.3, 6.4) were also achieved to an overall satisfactory extent.

Bibliography


Richards, J. (1971), "A Non-Contrastive Approach to Error Analysis". In


Appendix I
A sample of the 'Test Candidate Background Questionnaire' as used to investigate the linguistic education, experience and backgrounds of the members of the test group prior to the EA survey (see 5.3.6 and 5.3.5 (iv).

Test Candidate Background Questionnaire
Please tick the boxes (☑) and write answers where necessary. All information will be treated confidentially.

Part A: About You
Name:................................. Age: ....................... Sex: M ☐ F ☐
Previous occupation in Japan: Date of arrival in England:
..............................................................................................

Part B: About Studying English
1. When did you last study English?
☐ High School ☐ English Conversation School
☐ University ☐ Other: ........................................

2. Have you ever studied English abroad?
☐ Yes (If 'yes', where, when and for how long?):
☐ No............................................................................................................

3. Do you already have any English qualifications?
☐ TOEIC (Score: .................) ☐ Cambridge PET
☐ TOEFL (Score: .................) ☐ EIKEN (Level: ..................)
☐ Other: ................................................

4. Have you ever studied any other languages?
☐ Yes (If 'yes', please specify):
☐ No............................................................................................................

( 72 )
Part C: About Speaking English

5. Do you use English at work in Japan?
□ Yes (If 'yes', please specify):
□ No.............................................................................................................

6. Have you ever lived, worked or made a business trip to an English-speaking country?
□ Yes (If 'yes', please specify):
□ No.............................................................................................................

7. Have you ever travelled to an English-speaking country?
□ Yes (If 'yes', where, when and for how long?):
□ No.............................................................................................................

8. Do you have any native English-speaking friends?
□ Yes (If 'yes', which nationality):
□ No.............................................................................................................

9. Who do you speak English to regularly?
□ teacher(s) □ nobody
□ other students □ other: ........................................
□ host family

Thank you for your cooperation.
Appendix II
A sample of Part I of the test used during the EA survey, designed to investigate aural phonetic skills (see 5.3.3). Also below is a copy of the answer key used during marking (see 5.4.1).

Part I: Aural Test of Recognition and Discrimination of English Phonemes:

Instructions: Tick the box (✓) next to the word that you hear on the tape.
For example:

He’s always been something of a landlord.

1. He works as a
   □ warder.
   □ hoarder.
2. I think we need to hire another
   □ rover.
   □ robber.
3. That party on Saturday was such a
   □ ball.
   □ bore.
4. If we don’t issue a threat or
   □ two
   □ sue
   they’ll never change their minds.
5. I’m Tim and his name’s
   □ Jess.
   □ Jeff.
6. □ hate rhymes with ‘mate’.
   □ wait
7. This particular product requires a hard
   □ sell.
   □ shell.
8. I’ll take these pills and
   □ doze
   □ those
   over there.
9. I don’t think that’s my
   □ fault.
   □ salt.
10. Don't touch his ☐ wood or there'll be trouble.
   ☐ hood
11. She has a problem with her ☐ bowels.
   ☐ vowels.
12. It doesn't feel ☐ light to me.
   ☐ right
13. We've come across a few ☐ witches while driving in the countryside.
   ☐ hitches
14. They ☐ load the toilet paper at work.
   ☐ loathe
15. What a good ☐ theme.
   ☐ team.
16. Mind the ☐ curve when you're driving.
   ☐ curb
17. I can't take any more of this ☐ heat.
   ☐ wheat.
18. He's training to be a ☐ pirate.
   ☐ pilot.
19. It's best to ☐ shave as much as possible.
   ☐ save
20. You can't stop rabbits from ☐ breathing so quickly.
   ☐ breeding
<table>
<thead>
<tr>
<th>Test Item Number</th>
<th>Test Speech Sound or Sequence</th>
<th>Correct Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>[ wɔə ]</td>
<td>warder</td>
</tr>
<tr>
<td>2.</td>
<td>medial [ v ]</td>
<td>rover</td>
</tr>
<tr>
<td>3.</td>
<td>final [ ə ]</td>
<td>bore</td>
</tr>
<tr>
<td>4.</td>
<td>[ tʰuː ]</td>
<td>two</td>
</tr>
<tr>
<td>5.</td>
<td>final [ f ]</td>
<td>Jeff</td>
</tr>
<tr>
<td>6.</td>
<td>[ wɛi ]</td>
<td>wait</td>
</tr>
<tr>
<td>7.</td>
<td>[ ʃ ɛ ]</td>
<td>shell</td>
</tr>
<tr>
<td>8.</td>
<td>initial [ ɔ ]</td>
<td>those</td>
</tr>
<tr>
<td>9.</td>
<td>initial [ f ]</td>
<td>fault</td>
</tr>
<tr>
<td>10.</td>
<td>[ wʊ ]</td>
<td>wood</td>
</tr>
<tr>
<td>11.</td>
<td>initial [ v ]</td>
<td>vowels</td>
</tr>
<tr>
<td>12.</td>
<td>initial [ ɹ ]</td>
<td>right</td>
</tr>
<tr>
<td>13.</td>
<td>[ wi ]</td>
<td>witches</td>
</tr>
<tr>
<td>14.</td>
<td>final [ ɔ ]</td>
<td>loathe</td>
</tr>
<tr>
<td>15.</td>
<td>[ tʰi ]</td>
<td>team</td>
</tr>
<tr>
<td>16.</td>
<td>final [ v ]</td>
<td>curve</td>
</tr>
<tr>
<td>17.</td>
<td>[ wi ]</td>
<td>wheat</td>
</tr>
<tr>
<td>18.</td>
<td>medial [ ɹ ]</td>
<td>pirate</td>
</tr>
<tr>
<td>19.</td>
<td>[ ʃɛi ]</td>
<td>shave</td>
</tr>
<tr>
<td>20.</td>
<td>medial [ ɔ ]</td>
<td>breathing</td>
</tr>
</tbody>
</table>
Appendix III

A sample of Part II of the test used during the EA survey, designed to investigate oral phonetic skills (see 5.3.3). Also attached is a copy of the testee grade sheet used during marking (see 5.4.2).

Part II: Oral Test of Production of English Speech Sounds

Instructions: Form sentences about Tim using the pictures and words below. For example:

**Sentence:** "Tim likes listening to the radio."

About Tim

- **funny**
- **(has a) shave**
- **smooth (face)**

Tim's Street, House and Garden

- **(on a) curve**
- **right**
- **wood**
Tim's Job

- thief
- pirate

Tim's Hobbies

- war (films)
- (soccer) team
- two (dogs)
- rover

Tim's Dislikes

- wait
- vowels

Tim's Ambitions

- win
- then...
Testee Grade Sheet

Testee's Name: ............................................................

(Tick the appropriate grade ☑)

<table>
<thead>
<tr>
<th>Test Item Number</th>
<th>Test Speech Sound or Sequence</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>initial [ f ] (funny)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td>2.</td>
<td>[ʃɛt] (shave)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td>3.</td>
<td>final [ ə ] (smooth)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td>4.</td>
<td>final [ v ] (curve)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td>5.</td>
<td>initial [ r ] (right)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td>6.</td>
<td>[ wʊ ] (wood)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td>7.</td>
<td>final [ ə ] (fire)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td>8.</td>
<td>[ wi ] (weeds)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td>9.</td>
<td>medial [ ə ] (heather)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td>10.</td>
<td>[ʃɛd] (shed)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td>11.</td>
<td>final [ f ] (thief)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td>12.</td>
<td>medial [ ə ] (pirate)</td>
<td>☐ Satisfactory ☐ Poor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>13.</td>
<td>[wɔː ] (war)</td>
<td>□ Satisfactory □ Poor</td>
</tr>
<tr>
<td>14.</td>
<td>[tʰi ] (team)</td>
<td>□ Satisfactory □ Poor</td>
</tr>
<tr>
<td>15.</td>
<td>[tʰuə ] (two)</td>
<td>□ Satisfactory □ Poor</td>
</tr>
<tr>
<td>16.</td>
<td>medial [v ] (rover)</td>
<td>□ Satisfactory □ Poor</td>
</tr>
<tr>
<td>17.</td>
<td>[wəɪ ] (wait)</td>
<td>□ Satisfactory □ Poor</td>
</tr>
<tr>
<td>18.</td>
<td>initial [v ] (vowels)</td>
<td>□ Satisfactory □ Poor</td>
</tr>
<tr>
<td>19.</td>
<td>[wi ] (win)</td>
<td>□ Satisfactory □ Poor</td>
</tr>
<tr>
<td>20.</td>
<td>initial [θ ] (then)</td>
<td>□ Satisfactory □ Poor</td>
</tr>
</tbody>
</table>