Exploring measures and perceptions of fluency in the speech of second language learners

Judit Kormos and Mariann Dénes
Eötvös Loránd University, Budapest

Exploring measures and perceptions of fluency

ABSTRACT

The research reported in this paper explores which variables predict native and non-native speaking teachers' perception of fluency and distinguish fluent from non-fluent L2 learners. In addition to traditional measures of the quality of students' output such as accuracy and lexical diversity, we investigated speech samples collected from 16 Hungarian L2 learners at two distinct levels of proficiency with the help of computer technology. The two groups of students were compared and their temporal and linguistic measures were correlated with the fluency scores they received from three experienced native and three non-native speaker teacher judges. The teachers' written comments concerning the students' performance were also taken into consideration. For all the native and non-native teachers, speech rate, the mean length of utterance, phonation time ratio and the number of stressed words produced per minute were the best predictors of fluency scores. However, the raters differed as regards how much importance they attributed to accuracy, lexical diversity and the mean length of pauses. The number of filled and unfilled pauses and other disfluency phenomena were not found to influence perceptions of fluency.

INTRODUCTION

When one inquires about someone's level of proficiency, the answer is often that "I can speak the language fluently." Speaking a language fluently is frequently the ultimate goal to be attained in mastering a language. Despite the fact that the terms "fluency" and "fluently" are regularly used in language pedagogy and language testing as well as in various fields of applied linguistics, there seems to be no consensus concerning what is understood by these concepts (Chambers, 1997). Moreover, it is not only the definition of fluency that has been a matter of debate, but its measurement as well (see e.g. Koponen and Riggenbach, 2000). Research investigating what variables underlie listeners' perception of fluency has also been very scarce. The study of how fluency can be measured in a reliable way is important in several fields. In language testing the candidates' fluency is frequently judged, and by knowing which temporal measures of fluency contribute best to the listeners' perception of fluency and distinguish fluent and non-fluent speakers, more reliable criteria can be

developed. This knowledge is also useful in language pedagogy since it can help learners in enhancing their fluency. Establishing reliable measures of fluency is also important for researchers in applied linguistics since fluency of students' oral production is often assessed in a number of studies (e.g. task-based language learning research Skehan, 1996; Skehan and Foster, 1997; 1999).

Although there is a large body of pausological research (for a review see Griffiths, 1991), studies on perceptions of fluency are not so numerous. Inspired by Riggenbach's (1991) ground-breaking work, a few follow-up studies have been conducted to investigate which variables contribute to perceptions of fluency. Most of these studies, however, suffer from several methodological shortcomings partly because they use very few subjects and fail to employ reliable methods of analysing the duration of pauses (e.g. Ejzenberg, 2000; Freed, 1995; 2000; Lennon, 1990; Riggenbach, 1991). One of the large scale studies that investigated perceptions of fluency and produced reliable results was conducted by van Gelderen (1994), which, however, investigated L1 speakers of Dutch. Therefore it is not clear to what extent his results are applicable for L2 learners. In another important research project Fulcher (1996) analysed what kind of hesitation phenomena are able to predict fluency scores awarded in language proficiency exams. His study, however, failed to consider certain temporal variables such as speech rate and pauses shorter than 3 seconds, both of which might be important in the perceptions of fluency.

Although there are two types of definitions of fluency: one which considers fluency as a temporal phenomenon, and one that regards it as spoken language competence (for more details on definitions see the section below), fluency research suffers from the lack of studies that investigate a combination of linguistic, temporal, phonological and interactional variables. Thus we have limited knowledge what role other variables such as accuracy, lexical diversity, grammatical complexity and intonation play in perceptions of fluency. Moreover, information is completely lacking on whether native and non-native speakers show any differences in how they conceive of fluency.

The aim of our research is to investigate what linguistic and temporal variables predict native and non-native speaking teachers' perception of fluency. In this study teachers were selected as judges because it is most frequently the members of this profession who have to assess this aspect of speech production. In addition, they are the ones who make explicit judgements in various types testing situations and in deciding how to improve students' fluency. Analysing temporal features of spoken

texts is very time-consuming and difficult, therefore we could only investigate the speech samples of 16 learners of English as a foreign language, but this number is higher than in most earlier research and allows for the use of non-parametric statistics. We collected samples from two proficiency groups and analysed them with the help of a computer program, which is able to identify speech segments and silent pauses to a high degree of precision. The two groups were then compared along a wide range of temporal variables. We also correlated measures of accuracy, lexical richness and productivity with fluency scores awarded by three experienced native and non-native speaking teachers and testers in order to establish which of these variables are more influential in perceptions of fluency.

REVIEW OF LITERATURE

Definitions of fluency

In one of the first studies investigating fluency, Fillmore (1979) conceptualised fluency in four different ways. First, he defined fluency as the ability to talk at length with few pauses and to be able to fill the time with talk. Second, a fluent speaker is not only capable of talking without hesitations but of expressing his/her message in a coherent, reasoned and "semantically densed" manner. Third, a person is considered to be fluent if he/she knows what to say in a wide of range of contexts. Finally, Fillmore (1979) argued that fluent speakers are creative and imaginative in their language use and a maximally fluent speaker has all of the above mentioned abilities. Fillmore's definition of fluency is very extensive, but it is unclear how this conceptualization differs from the definition of global oral proficiency.

One of the first definitions of second language fluency was provided by Pawley and Syder (1983), who regard native-like fluency as "the native speaker's ability to produce fluent stretches of discourse" (p. 191). This definition is of much narrower scope than that of Fillmore and has served as a basis for several further studies.

Lennon (1990, 2000) pointed out that fluency is usually used in two senses. In the so-called broad sense, fluency seems to mean global oral proficiency, that is, a fluent speaker has a high

command of the foreign or second language. The definition proposed by Sajavaara (1987) can also be regarded as a broad conceptualisation of fluency. He defined fluency as "the communicative acceptability of the speech act, or 'communicative fit'" (p. 62). He also points out that expectations concerning what is appropriate in a communicative context vary according to the situation, therefore his definition seems to be very difficult to operationalise. This conceptualisation of fluency bears resemblance to the third aspect of fluency described by Fillmore (1979).

In its narrower sense, fluency can be considered one component of oral proficiency, which is often used as one of the scores in assessing candidates' oral language skills in an exam situation. Lennon (1990) pointed out that fluency differs from the other scores in oral language exams (e.g. accuracy, appropriacy, etc.) in that it is purely a performance phenomenon, and consequently defined fluency as "an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently" (p. 391). Thus he argued that "fluency reflects the speaker's ability to focus the listener's attention on his/her message by presenting a finished product, rather than inviting the listener to focus on the working of the production mechanisms" (pp. 391-392). Rehbein (1987) provided a similar definition, claiming that "fluency means that the activities of planning and uttering can be executed nearly simultaneously by the speaker of the language" (p. 104). He also added that fluency depends on the context, namely on the "speaker's evaluation of the hearer's expectations" (p.104). Schmidt (1992) refined Lennon's (1990) definition by adding that fluency in speech production is "automatic procedural skill" (based on Carlson, Sullivan, & Schneider, 1989) and that fluent speech "is automatic, not requiring much attention or effort" (Schmidt, 1992, p. 358). In a more recent study, Lennon (2000) synthesised earlier definitions and proposed that "a working definition of fluency might be the rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing (p. 26)."

Measures of fluency

Just as defining fluency is rather problematical, the establishment of the components of fluency is not without difficulty, either. Four different approaches to delineating the measures of fluency exist

in the investigation of L2 learner's speech. The first trend of research is concerned with the temporal aspects of speech production (e.g. Lennon 1990; Möhle, 1984), the second combines these variables with the investigation of interactive features (e.g. Riggenbach, 1991) and the third approach explores the phonological aspects of fluency (e.g. Hieke, 1984, Wennerstrom, 2000) as well. Finally, recent studies have included the analysis of formulaic speech in studying fluency in second language speech (e.g Ejzenberg, 2000; Towell et al., 1996).

A number of studies have been concerned with establishing the appropriate measures of fluency. The empirical studies in this field used three different approaches: they either investigated the development of fluency longitudinally (Freed, 1995; 2000; Lennon, 1990; Towell et al.,1996), or compared fluent and non-fluent speakers (Ejzenberg, 2000; Riggenbach, 1991; Tonkyn, 2001) or correlated fluency scores with temporal variables (Rekart and Dunkel, 1992, Fulcher, 1996). We have to note again that the number of participants investigated was very small in most of these research projects, and in many of them no statistical analyses and computer technology for identifying pauses reliably were used. Nevertheless most of them conclude that the best predictors of fluency are *speech rate*, that is, the number of syllables articulated per minute and the *mean length of runs*, that is, the average number of syllables produced in utterances between pauses of 0.25 seconds and above (e.g. Ejzenberg, 2000; Freed, 1995, 2000; Lennon, 1990; Riggenbach, 1991, Towell et al, 1996). *Phonation-time ratio*, that is, the percentage of time spent speaking as a percentage proportion of the time taken to produce the speech sample, was also found to be a good predictor of fluency (Towell et al, 1996; Lennon, 1990; van Gelderen, 1994). Research findings are equivocal concerning the frequency of filled and unfilled pauses as well as disfluencies such as repetitions, restarts and repairs. The studies with small number of participants found that the frequency of silent and filled pauses distinguished between fluent and non-fluent speakers (e.g. Freed, 1995, 2000; Lennon, 1990; Riggenbach, 1991). On the other hand, in research projects in which a higher number of students participated, the number of filled and unfilled pauses and ratings of fluency did not correlate (Rekart and Dunkel, 1992; van Gelderen, 1994). Most researchers agree that disfluencies tend to occur in clusters in the speech of non-fluent L2 learners (e.g. Freed, 1995, 2000; Riggenbach 1991), while fluent students tend to pause at grammatical junctures (Lennon, 1990; Towell et al., 1996). Fulcher (1996) looked at how the reasons why students hesitate change with the development of L2

competence. He found that low-proficiency students tend to hesitate because they have problems retrieving lexical items, encoding the grammatical form of their message and correcting their own output. On the other hand, high-proficiency students are able to plan in advance and mostly hesitate only when they want to express complex ideas.

Based on the assumption that fluency is context-dependent (e.g. Rehbein, 1987; Sajavaara, 1987; Lennon, 1990), Riggenbach (1991) complemented the analysis of temporal variables underlying second language fluency with the investigation of interactive features. Her results revealed that topic-initiations, backchannels, substantive comments, latching and overlapping as well as the amount of speech produced also contributed to fluency judgements, though to a limited extent.

In the field of phonological research, Hieke (1985) established additional measures of fluency on the basis of the assumption that fluent speech equals connected speech, in which certain phonological procedures, such as *consonant attraction* are at work. Consonant attraction "occurs where final consonants are drawn to the following syllable if that begins with a vowel" (Hieke, 1985, p. 140). In an earlier study, Hieke (1984) found that consonant attraction can be a reliable indicator of the fluency of non-native speech in informal English style. Wennerstorm (2000) in her research investigated in what ways intonation influences the perception of fluency by means of analysing dialogues between speakers of English as a second language and native English speakers. Her study suggests that it is the ability to speak in phrases instead of speaking word by word that can lead to the perception of fluent speech, rather than longer utterances or shorter pauses. In another study, Vanderplank (1993) suggests that pacing (the number of stressed words per minute) and spacing (the proportion of stressed words to the total number of words) are better indicators of difficulty in listening materials than standard speech rate measures such as syllable per minute. Indirectly, this would mean that these variables are also useful in predicting fluency scores.

Towell et al (1996) investigated what qualitative changes take place in the use of formulaic language parallel to the increase of fluency after participants spent a year in the target language environment. They found that the two selected students improved in how they employed different types of formulae after their stay abroad. Ejzenberg (2000) compared how fluent and non-fluent speakers employ formulaic language. Her results also showed that fluent students were able to make

use of prefabricated chunks more efficiently, whereas non-fluent learners frequently used formulae inappropriately.

METHOD

Participants

In order to investigate the differences between fluent and non-fluent speakers, two distinct groups of learners were selected. The eight participants of the advanced group were drawn from Hungarian students enrolled in the School of English and American Studies, Eötvös Loránd University, Budapest, in their third, fourth or fifth years of studies. Participants in this group were all females, whose age ranged from 19 to 30. As regards the participants' language learning history, they all had learnt English in a formal setting both at secondary school and/or at language schools for a minimum of five years, and they also spent a period of 6 – 12 months in an English speaking country. All of the participants in this group had high scores at the language proficiency exam administered at the university.

Participants of the low-intermediate proficiency group were selected from Hungarian students attending a language school where they all learnt English in the same group. The students were enrolled in their groups based on their scores in a placement test, with the help of which it could be ensured that their level of proficiency was similar. Like the participants in the advanced group, their age ranged between 19 and 30. Six of them were females and two of them were males. They had a similar language learning history in the sense that they all had learnt English at secondary school, and none of them had been to an English speaking country.

In this research the three non-native speaking judges of the participants' fluency were Hungarian native speakers who were experienced university teachers of applied linguistics and examiners in the language proficiency exams administered at the school. The three teachers were all females with at least 10 years of experience in the profession. In selecting the native speaking participants we wanted one standard British, one standard American speaker and a speaker with a non-standard British accent to be

represented. Thus a male speaker of standard British English, and a male teacher, who was a speaker of English with a slight Scottish accent were selected. Both of these teachers had at least 10 years experience in teaching and also acted as examiners at the university. The third native speaker participant was a young female American visiting teacher, with a few years experience in teaching but with limited experience as a language tester.

All the participants, both the teachers and the students were informed of the purpose of the research, and they participated in it on a voluntary basis. Participants' anonymity was preserved in every phase of the study.

Procedures

Materials collected for this research involved tape-recorded speech samples that were 2-3 minutes long on average. Participants carried out a narrative task which was similar to traditionally used elicitation devices in pausological research (e.g. Riazantseva, 2001). The students were asked to choose from three sets of cartoon strips and were instructed to make up a story related to the selected cartoon. The cartoons used in the description task consisted of a series of 6-10 pictures arranged in a logical order. Cartoons were taken from popular English course books and were designed for similar tasks. Selection criteria for the cartoons included relative simplicity of the story and of the vocabulary necessary to describe it (Riazantseva, 20001). The choice of narrative was motivated by two reasons. First, computer-analysis of speech phenomena becomes very difficult in an interactive task, where it might happen that the two speakers talk simultaneously. Second, it is known from very early research on pausology (Goldman-Eisler, 1968) and also from task-based research (for a review see Skehan, 1998) that having to produce different types of content places different cognitive load on speakers, which, in turn, influences the fluency of the production. By providing fixed content (a series of pictures that form a story), the influencing factor of content could be eliminated. Time devoted to planning was specified in two

minutes. These two minutes were provided for the participants to have enough time to understand the story depicted in the cartoon and to gather their thoughts about how they will narrate it. One of the researchers was always present at the recording but did not intervene in case of hesitations or problems. The students' performance was recorded on an audio-tape.

The three native and non-native speaking assessors judged the oral performances of the participants on a semantic differential scale that ranged from one to five, where one corresponded to the least fluent and five to the most fluent speech. The judges did not know the participants of the research and the speech samples were mixed on the tape. Descriptors for the five categories were not given in order to gain intuitive judgements, but they were asked to comment on their scores for each participant.

Analysis

In order to gain precise temporal measures, the speech samples were transcribed with the help of a computer program called Transcriber. By means of this program each silent pause was detected and measured in milliseconds. Every speech run distinguishable from the amplitude of the background noise was transcribed by the researcher, and with the help of the program the duration of speech runs was also measured. In this study we examined the following ten temporal variables:

1. Speech rate

   Rate of speech was calculated according to the method recommended by Riggenbach (1991). The total number of syllables produced in a given speech sample was divided by the amount of total time required to produce the speech sample, (including pause time) expressed in seconds. This figure was then multiplied by sixty to give a figure expressed in syllables per minute. Following Riggenbach's suggestions (1991), unfilled pauses under 3 seconds were not included in the calculation of speech rate. Unfilled pauses shorter than 3 seconds are generally regarded articulation pauses and not as hesitation phenomena.

2. Articulation rate

In calculating the articulation rate the total number of syllables produced in a given speech sample was divided by the amount of time taken to produce them in seconds, which was then multiplied by sixty. Unlike in the calculation of speech rate, pause time was excluded. Articulation rate is expressed as the mean number of syllables produced per minute over the total amount of time spent speaking when producing the speech sample. Following Riggenbach (1991), in the articulation rate all semantic units were counted, "including filled pauses and partial words (using the criterion that partial words contain not just an initial consonant but also a vowel and thus are recognizable as words)" (p.428).

3. Phonation-time ratio

Phonation-time ratio was calculated as the percentage of time spent speaking as a percentage proportion of the time taken to produce the speech sample" (Towell et al., 1996, p. 91).

4. Mean length of runs

The mean length of runs was calculated as an average number of syllables produced in utterances between pauses of 0.25 seconds and above. Towell et al. (1996) point out that there has been an ongoing debate among researchers about the cut-off point of pause length. If this point is too low, the stop phase indicated by the computer may signal a plosive or other phenomena that should not be considered as hesitations. If the cut-off point is too high, some amount of time may be omitted from the analysis. Therefore Towell et al. argue that pauses above 0.25 seconds are the most reliable cut-off points.

5. The number of silent pauses per minute

In analysing pauses, pauses over 0.2 seconds were considered. Pauses shorter than 0.2 seconds are considered micro-pauses and are not regarded as hesitation phenomena (Riggenbach, 1991). Due to the fact that the speech samples were not of equal length, the total number of pauses was

divided by the total amount of time spent speaking expressed in seconds and was multiplied by 60.

6. The mean length of pauses

The mean length of pauses was calculated by dividing the total length of pauses above 0.2 seconds by the total number of pauses above 0.2 seconds.

7. The number of filled pauses per minute

The total number of filled pauses such as uhm, er, mm were divided by the total amount of time expressed in seconds and was multiplied by 60.

8. The number of disfluencies per minute

The total number of disfluencies such as repetitions, restarts and repairs were divided by the total amount of time expressed in seconds and was multiplied by 60.

9. Pace

The number of stressed words per minute (Vanderplank, 1993).

10. Space

The proportion of stressed words to the total number of words (Vanderplank, 1993).

The quantity of talk students produced was measured by the total number of words (see also Dewaele, 2000; Dewaele and Pavlenko, 2003; Dörnyei and Kormos, 2000). In order to assess lexical diversity, we used Malvern's and Richards' (1997) *D-formula*. This formula is based on the widely used measure of lexical variety, the type-token ratio (TTR), which is the total number of different words (types) divided by the total number of words (tokens) produced. The problem with the TTR is, however, that it depends on the sample size, that is, on the number of words spoken by the participants. Richards (1987) found that the "type-token ratio falls rapidly as the number of tokens increases" (p. 205). In a recent study Jarvis (2002) argued that Malvern's and Richards' (1997) D-formula can be used to measure lexical richness in L2 texts in a reliable way, therefore we applied this formula to establish a measure of lexical diversity. [1] Accuracy was measured by *the proportion of error-free clauses relative to the total number of clauses*. This measure has been widely

[1] The calculation of the D-value is based on a mathematical probabilistic model, and the software available at the CHILDES web-site: http//childes.psy.cmu.edu uses random sampling of tokens in plotting the curve of TTR against increasing token size for the text to be investigated.

used in task-based research and has been proven to reflect the accuracy grammatical complexity of students' output in a reliable manner (see Bygate, 1999; Foster, and Skehan, 1996; Skehan and Foster, 1997).

We computed Spearman rank-order correlations between the temporal and linguistic variables and students' fluency scores and also looked at the correlations between the variables. Fluent and non-fluent speakers were compared by means of the Mann-Whitney U-test.

RESULTS

First of all, we looked at the comments of the assessors in order to establish what factors influenced their judgements. The non-native speaker teachers most frequently they justified the awarded scores with reference to the naturalness and ease of flow of speech, the presence or lack of pauses and self-corrections, the speed of delivery and the efficiency with which students handle breakdowns. Other important reasons behind their judgements were the accuracy of the output and the range of vocabulary employed. Sometimes they mentioned that the participant was not verbose or that they used linking words appropriately. The native speaker teachers also gave varied reasons for their judgements. All of them referred to speed of delivery, the absence or abundance of hesitation phenomena, but they differed as regards the importance of lexical variety and accuracy. While all the comments of the American teacher mentioned the accuracy of output, and the standard British English speaker also made frequent reference to grammatical and lexical errors, the Scottish teacher did not mention this aspect of performance. The inter-rater reliability for the three non-native speaker teachers was 0.78, which considering the fact that no descriptors were given can be regarded as acceptable. The inter-rater reliability for the native speakers was considerably lower (r= 0.72).

As can be seen in Table 1, the Mann-Whitney U-test revealed significant differences between fluent and non-fluent participants in the case of five of the investigated ten temporal variables: speech

---

Malvern and Richards (1997) argue that the D-value is a valid measure of diversity because it does not depend on the length of the sample, and it uses all the words produced by the participants.

rate, phonation time ratio, the mean length of runs and of pauses as well as pace (the number of stressed words per minute). Advanced students spoke faster, silent pauses took up a smaller proportion of their speaking time, produced longer stretches of discourse between pauses, used shorter pauses and uttered more stressed words within a minute than low-intermediate students. Fluent speakers produced more accurate and lexically more diverse output than their less fluent counterparts.

Insert Table 1 around here

We calculated two composite fluency scores for the participants: one of them was the sum of scores awarded by the native speaker teachers and the other the sum of the non-native speaker scores. Table 2 shows the rank-order correlations with these two types of composite scores and with each of the raters' scores. The results indicate that there is a set of variables that can predict both the composite and the individual raters' fluency scores in a reliable way. These are the speech rate, the phonation time ratio, the mean length of runs and the number of stressed words per minute (pace). Each of these temporal measures correlates strongly both with the native and the non-native teachers' scores. Among them speech rate, the mean length of runs and pace are the best predictors, which can account between 60 and 80% of the variance in the fluency scores. The mean length of pauses is also significantly related to the composite native and non-native fluency scores, but it does not correlate with all of the individual raters' scores. It seems that for one of the native and one of the non-native teachers it was an important factor that influenced their judgements, but the rest of the assessors did not take the length of pauses into consideration. The importance of linguistic measures in perceptions of fluency is also ambiguous. While the native and non-native teachers' composite scores show moderate correlations with accuracy, lexical variety and productivity, individual raters differ as regards these measures. One of the native and one of

the non-native teachers did not attribute great importance to accuracy. One of the non-native teachers (NNS3) consistently did not take any of the linguistic measures into consideration. Table 2 also shows that certain temporal variables such as articulation rate, the number of silent and filled pauses and the number of disfluencies per minute as well as the ratio of stressed and unstressed words (space) are not related to fluency scores.

The rank-order correlations between the temporal variables shown in Table 3 reveal that pace, speech rate, phonation-time ratio, the mean length of runs and of pauses are strongly correlated with each other. The other cluster of temporal variables that are related include the number of filled and unfilled pauses and disfluencies per minute. The linguistic variables all significantly correlate with pace, speech rate and the mean length of runs with accuracy showing the highest correlation with these temporal measures.

Insert Table 2 around here

If we examine individual participant's scores and their performance on various temporal and linguistic measures we can see that Participant 15, who together with Participant 16 received the highest number of points, produced the highest number of stressed words per minute and her output was also the most accurate in the examined population. Participant 16 had the highest phonation time ratio and the mean length of pauses in her speech was the shortest. Participant 13 and 14, who both scored very high, were the fastest speakers in terms of syllables per minute and the mean length of their runs was also the longest. Their rank order according to the number of stressed words per minute is the third and fourth. If we look at the participants at the other end of the fluency scale, we can also see that those students who were judged to be the least fluent were the ones who produced the lowest number of stressed words per minute (e.g. Participant 1, 2 and 3). Participant 2, however, had relatively higher speech rate and uttered

longer units between pauses, which is why one of the raters (NS3) perceived her to be more fluent. Participant 6 is an interesting case because she is the second slowest speaker in terms of speech rate, her phonation time ratio is also the second lowest and her pauses tend to be very long. Nevertheless she was still perceived to be more fluent by two of the non-native speaker judges than Participant 2, who spoke almost twice as fast because her accuracy was very good (86% of her clauses did not contain any mistake).

Insert Table 3 around here

DISCUSSION

The findings of this study indicate that there is a set of variables that are very good predictors of fluency scores both for native and non-native speaker judges: speech rate, the mean length of runs, phonation-time ratio and pace. The results concerning speech rate, the mean length of runs and phonation-time ratio are not new, since other researchers (e.g. e.g. Ejzenberg, 2000; Freed, 1995, 2000; Lennon, 1990; Riggenbach, 1991; Towell et al, 1996, van Gelderen, 1994) have also argued for the use of this measure. The finding that pace, which is a temporal variable that also considers one specific feature of intonation, namely stress, is an equally reliable predictor of fluency judgements, is novel because Vanderplank (1993) only investigated the role of pace in judgements of listening difficulty. The results are striking because one can rarely see such high correlation in studies in applied linguistics especially if the number of subjects is relatively low. If we look at Table 3, it can be seen that with a few exceptions, the rank order that can be set up based on pace is almost the same as the rank order of fluency scores. This and the rate of correlations seems to be a very strong evidence for the importance of this relatively underused variable in fluency research. Thus we can argue that how many stressed words one can say in a minute is a slightly better predictor of fluency than how many syllables one utters a minute. In other words if a speaker utters a lot of unstressed words with a high speed, he or she is not necessarily perceived to be very fluent. Our finding suggests that if one wants to make an informed judgement of someone's fluency either in a language test or in an empirical study, the measure of pace could also be used. This variable is relatively easy to calculate and seems to be an excellent predictor. As Vanderplank (1993) argues, pace can also be useful in grading the difficulty of listening passages as the more fluent a speaker is, the more difficulties an L2 learner might have in understanding his or her message.

The finding that a high proportion of the variation in fluency scores can be explained with a cluster of temporal variables also shows that fluency is primarily a temporal and intonational

phenomenon. As regards the frequency of filled and unfilled pauses and disfluencies, our study lends support to the work of those researchers who claim that these variables do not affect perceptions of fluency (Rekart and Dunkel, 1992; van Gelderen, 1994). The results of our study also show that there are two clusters of temporal variables; one which includes pace, the mean length of runs and pauses, speech rate and phonation-time ration and the other that is comprised of the frequency of filled and unfilled pauses and other disfluencies. The former group of variables that is related to the speed of delivery seems to be an important factor in fluency judgements, while the latter is not.

This research has also lent support to Fillmore's (1979) mainly intuitive claim that perceptions of fluency are related to productivity as we found moderate correlations between the number of words produced and participants' fluency scores. It seems that also in L2, talkative speakers are regarded as more fluent.

Our study shows that fluency is not only a temporal phenomenon, as raters do not only look at speed and pace when intuitively judging someone's fluency, but consider other variables strongly related to proficiency such as accuracy and lexical diversity. Our results support earlier theoretical conceptualisations of fluency according to which there exist two senses of fluency: low-order fluency (temporal aspects of fluency) and high-order fluency that can be equated with proficiency (e.g. Lennon, 1990, 2000). Our research suggests that accuracy also plays an important role in fluency judgements and sometimes overrides the effect of temporal factors on listeners (e.g. in the case of Participant 6). The correlations between the temporal and linguistic variables also reveal that accuracy is positively related to temporal variables that are influential in fluency judgements. In other words, it seems that those students who were fluent in terms of speed and pace also produced accurate output. In psycholinguistic terms this means that one is only able to speak fluently if speech production mechanisms are largely automatic and if automatic sequences are memorised, retrieved and used accurately (see Schmidt, 1992 for a review). We also have to note that exceptions to this also exist such as Participant 6, whose production was very accurate but slow. Low-proficiency students generally cannot rely on a sufficient

number of automatic sequences and apply conscious rule-based mechanisms, and if they strive to be highly accurate, their speech becomes very slow. Thus in certain cases especially among less competent speakers, speed and accuracy might be in inverse relationship with each other. We have to note, however, that just as researchers are divided on the definitions of fluency, teachers also differ as regards the conceptualisation of this phenomenon. For some of them, it includes accuracy and lexical diversity, while some other teachers disregard these aspects of performance.

On theoretical grounds our study indicates that fluency as a concept needs to include speed, pace, smoothness and accuracy. Thus, if the aim of language teaching is to help students to become fluent speakers, both exercises that prompt students to express their communicative intent within the limited time-constraints of real-life interactions and those that promote accuracy should be applied. Moreover, tasks that develop accuracy and fluency in the temporal sense at the same time seem to be the most useful.

For research and language testing purposes, however, it is often necessary to distinguish between temporal fluency and accuracy. For researchers the conclusion of our study is that there is a new quick and easy way of establishing fluency with the help of the measure of pace, that is, the number of stressed words uttered per minute. In terms of language testing, we suggest that descriptors of fluency should include the speed of delivery, pace and the length of fluent units. Definitions of bands that include descriptors such as a high number of hesitations (filled and unfilled pauses) do not seem to be valid, as the raw number of these variables does not seem to be related to proficiency or fluency. This might be partly due to the fact that in certain situations native speakers also frequently hesitate and that individual speaking style might also influence hesitation behaviour. In addition, Fulcher (1993) argues that low and high-proficiency students might produce the same number of hesitations, but hesitate for different reasons and therefore create different impressions on listeners.

The results of our study also indicate that the investigated native and non-native teachers do not differ substantially in how they perceive fluency. Both native and non-native teachers base their

judgements on the same set of temporal variables. Despite the fact that individual differences exist between raters as regards accuracy and lexical diversity, these do not seem to be determined by native or non-native speaker status.

CONCLUSION

The study reported in this paper investigated differences between fluent and non-fluent L2 learners as well as the relationship of native and non-native teachers' perceptions of fluency and temporal and linguistic variables. In the research we analysed speech samples collected from 16 Hungarian L2 learners at two distinct levels of proficiency with the help of computer technology and used non-parametric statistical analyses. The results indicate that fluency is best conceived of as fast, smooth and accurate performance. If we regard fluency as a temporal phenomenon, it can also be characterised by pace (the number of stressed words per minute). The mean length of runs and speech rate were also found to be good indicators, but we also recommend the use of pace for measuring temporal fluency as it also includes one specific feature of intonation, namely stress, and it is easy to calculate. Phonation-time ratio and the mean length of pauses were also related to fluency scores, but this relationship was weaker than in the case of the mean length of runs and the speech rate. The number of filled and unfilled pauses and other disfluency phenomena were not found to influence perceptions of fluency. Our research also indicates that the accuracy of output plays an important role in fluency judgements and that accuracy and speed of delivery are positively related. In other words, fluent performance entails the application of efficient and accurate processing mechanisms.

Although the number of participants was higher than in most previous research, the limitation of this study is its small sample size. In order to support the results presented here, this research should be repeated with other types of tasks and with more L2 learners and raters.

REFERENCES

Bygate, M. (1999). Quality of language and purpose of task: Patterns of learners' language on two oral communication tasks. *Language Teaching Research, 3*, 185-214.

Carlson, R. A., Sullivan, M., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 517-526.

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

Dewaele, J.-M. (2000). Saisir l'insaisissable? Les mesures de longueur d'énoncés en linguistique appliquée. *International Review of Applied Linguistics,* 38, 31-47.

Dewaele, J.-M., & Pavlenko, A. (2003). Productivity and lexical diversity in native and non-native speech: A study of cross-cultural effects. In V. Cook (Ed.), *The effects of the second language on the first* (pp. 120-141). Clevedon: Multilingual Matters.

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research, 4,* 275-300.

Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 287-314). Michigan: The University of Michigan Press.

Fillmore, C. J. (1979). On fluency. In D. Kempler, and W. S. Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85-102). New York: Academic Press.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299-323.

Freed, B. (1995). What makes us think that students who study abroad become fluent? In B.Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123-48). Amsterdam: John Benjamins.

Freed, B. F. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243-265). Michigan: The University of Michigan Press.

Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing, 13,* 208 - 238.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* New York: Academic Press.

Griffiths, R. (1991). Pausological research in an L2 context: A rationale, and review of selected Studies. *Applied Linguistics 12,* 345 - 364.

Hieke, A. E. (1984). Linking as a marker of fluent speech. *Language and Speech, 27*, 343-354.

Hieke, A. E. (1985). A componential approach to oral fluency evaluation. *Modern Language Journal, 69,* 135-142.

Jarvis, S. (2002). Short texts, best-fitting curves, and new measures of lexical diversity. *Language Testing 19*, 57-84.

Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 5-24). Michigan: The University of Michigan Press.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40,* 387-412.

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

Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language*. Clevedon: Multilingual Matters.

Möhle, D. (1984). A comparison of the second language speech production of different native speakers. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language production* (pp. 26-49). Tübingen: Günter Narr.

Pawley, A, & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 317-331). Amsterdam: Elsevier.

Rehbein, J. (1987). On fluency in second language speech. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 97-105). Norwood, N.J.: Ablex.

Rekart, D., & Dunkel, P. (1992). The utility of objective (computer) measures of the fluency of speakers of English as a second language. *Applied Language Learning, 3*, 65-85.

Riazantseva, A. (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition, 23,* 497-526.

Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of Child Language, 14,* 201-209.

Riggenbach, H. (1991). Towards an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes, 14,* 423-441.

Sajavaara, K. (1987). Second language speech production: Factors affecting fluency. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 45-65). Norwood, N.J.: Ablex.

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

Segalowitz, N. (2000). Automaticity and attention skill in fluent performance. In H. Riggenbach (Ed.), p*erspectives on fluency* (pp. 200-219). Michigan: The University of Michigan Press.

Skehan, P. (1996). A framework for the implementation of task based instruction. *Applied Linguistics, 17*, 38-62.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*, 185-211.

Skehan, P. & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*, 93-120.

Tonkyn, A. (2001). *The many voices of fluency*. Paper presented at the BAAL Annual Meeting, September 2001.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17,* 84-119.

Vanderplank, R. (1993). Pacing and spacing as predictors of difficulty in speaking and understanding English. *English Language Teaching Journal*, *47*, 117-125.

van Gelderen, A. (1994). Prediction of global ratings of fluency and delivery in narrative discourse by linguistic and phonetic measures - oral performances of students aged 11-12 years. *Language Testing, 11*, 291 - 319.

Wennerstorm, A. (2000). The role of intonation in second language fluency. In H. Riggenbach (Ed.). p*erspectives on fluency* (pp. 102-127). Michigan: The University of Michigan Press.

APPENDIX Examples for texts produced by the participants

Participant 16 (scored 15 points out of 15 in the fluency test) - length of pauses is indicated in brackets

So err (0.216) I would like to talk about this cartoon because it's quite er close to me (0.498) and this says that er I used to have four cats and we also went on holiday (0.140) with my family (0.652) and er taking that there was no one (0.190) err to look after the cats we had to take them with us (0.588). The problem was that these cats were used to living in flats (0.583) and when they saw (0.485) trees (0.499) they climbed up (0.138) they couldn't resist the temptation however they couldn't come down (0.660) so the problem arouse (0.1004) err (0.119) how to (0.429) you know take the cat off the (208) tree (0.772) so (0.535) this story is the following (0.557) there was a nice lady working in the garden (0.599) with his pussy cat (0.467) however (0.277) the cat (0.373) jumps up the tree (0.1214) the lady (0.166) is calling the cat (0.363) puss puss puss puss (0.1542) but the cat (0.427) does not respond (0.526) ok (0.816) the lady calls the err fire engine (0.866) she's shouting (0.319) help me help me (0.678) my cat is up the tree (0.548) ok we're coming (0.235) don't panic (0.312) we're coming (1.232) then (0.316) everything turns (0.894) quite good (0.891) the fire brigade arrives (0.693) and the cat (0.383) is taken down (0.258) the tree (1.870) the nice lady and the fire brigade are talking (0.377) they are having a tea (0.660) the atmosphere is relaxed (0.702) but suddenly (0.604) when they leave (0.598) the cat is run over (0.869) sad.

Participant 1 (scored 6 points out of 15 in the fluency test) - length of pauses is indicated in brackets

There was a lady (0.830) who (0.584) takes care (1.048) for his (0.743) his (0.158) her err garden (1.134) and she (0.141) has a (0.668) cat (0.781) and the cat err (0.406) climbed (0.342) err climb up (0.246) to a tree (1.914) the er (1.134) woman (0.924) is a little bit frustrated (0.933) cause err err (1.715) er (0.934) I think she s (0.390) says (0.851) the cat (0.920)don't c (0.233) err (0.027) doesn't care about it (2.268) she err (1.008) calls (1.013) the fire station (1.206) and then err (2.515) err (1.460) she tries (0.131) to (0.769) tries to err (3.361) call (0.628) it (0.383) down (1.047) with a (0.349) fish (2.098) but (0.985)nothing happens (0.575) the firemen came (0.677) and er (1.948)one of the firem (0.238) men (2.721)err (1.159) take (0.444) takes the (0.495) cat (0.254) down (2.646) then (0.235) everything is (0.840) going (0.927) well (0.731) the cat (5.752) the cat is ok (1.787) the (0.238) woman (0.303) is happy (2.283) the fire (0.357) men (2.026) err (2.305) are (0.644) invited for a (0.657) tea (1.973) and the err (1.523) happens something strange (2.484) and err (1.379) the cat (0.324) err (2.231) is (2.766) dead by a car (4.074) err (0.620) the firemen (1.625) go (2.708) go through the cat (1.550) with er (0.127) their (0.449) car.

ACKNOWLEDGEMENTS

We would like to thank the six participating teachers for the time and energy they devoted to rating. students' performance. We are grateful for the anonymous reviewers for their particularly thorough and

insightful comments. Thanks are also due to our colleague, Edit Hegybíró Kontra for her valuable critical remarks concerning the study and the paper.

Table 1. The comparison of the temporal and linguistic measures of low-intermediate and advanced students

Variable	Group	N	Mean	Sd.	Z	p
Speech rate	Low-intermediate	8	115.87	27.52	-3.04	0.001
	Advanced	8	181.19	30.42		
Articulation rate	Low-intermediate	8	227.45	27.94	-0.94	0.34
	Advanced	8	241.99	53.99		
Phonation time ratio	Low-intermediate	8	51.60	14.49	-2.31	0.02
	Advanced	8	69.05	6.45		
Mean length of runs	Low-intermediate	8	3.49	0.70	-3.36	0.001
	Advanced	8	6.23	1.15		
Number of silent	Low-intermediate	8	31.18	3.35	-.31	0.75
pauses per minute	Advanced	8	30.25	4.19		
Mean length of pauses	Low-intermediate	8	959.46	371.10	-1.99	0.04
in milliseconds	Advanced	8	615.08	121.91		
Number of filled	Low-intermediate	8	16.30	13.16	-.094	0.34
pauses per minute	Advanced	8	8.28	5.45		
Number of disfluencies	Low-intermediate	8	5.00	2.48	-0.84	0.44
per minute	Advanced	8	4.51	4.12		
Number of stressed	Low-intermediate	8	16.45	1.81	-3.36	0.001
words per minute	Advanced	8	33.12	8.1		
Ratio of stressed words	Low-intermediate	8	0.23	0.02	-0.63	0.52
and total no. of words	Advanced	8	0.23	0.06		
Number of error-free	Low-intermediate	8	0.53	0.19	-2.94	0.002
clauses/clauses	Advanced	8	0.86	0.05		
D-value	Low-intermediate	8	35.84	8.05	-2.94	0.002
	Advanced	8	57.39	11.34		
Number of words	Low-intermediate	8	159.37	49.43	-1.78	0.07
	Advanced	8	236.00	82.69		

Table 2. Rank order correlations of raters' scores and students' performance measures

	Speech rate	Art. rate	Phon.time ratio	Mean length of runs	No. of silent pauses/min	Mean length of pauses	No. of filled pauses /min	No. of disfluencies /min	No. of stressed words /min	Ratio of words and stressed words	Accuracy	D-value	No of words
NS composite score	.87**	.31	.80**	.91**	-.10	-.58*	-.08	-.10	.88**	-.06	.63**	.63**	.54*
NNS composite score	.81**	.36	.74**	.88**	-.09	-.62**	-.16	-.13	.92**	-.09	.76**	.57*	.56*
NS1 score	.77**	.15	.71**	.81**	-.09	-.60*	-.26	-.09	.88**	.12	.63**	.76**	.45
NS2 score	.74**	.17	.63**	.77**	-.03	-.45	-.07	-.08	.78**	-.20	.47	.54*	.54*
NS3 score	.86**	.50	.70**	.88**	-.28	-.49	-.20	-.20	.76**	-.01	.74**	.56*	.56*
NNS1 score	.68**	.23	.51*	.75**	-.10	-.38	-.23	-.12	.79**	-.01	.81**	.65**	.51*
NNS2 score	.82**	.45	.74**	.86**	-.28	-.49	-.11	-.18	.75**	-.04	.60*	.47	.51*
NNS3 score	.64**	.22	.71**	.71**	.06	-.70**	-.05	-.06	.84**	.19	.43	.42	.39

* indicates $p < 0.05$
** indicates $p < 0.01$

Table 3. Rank-order correlations of the temporal and linguistic variables

	Speech rate	Art. rate	Phon.time ratio	Mean length of runs	No. of silent pauses/min	Mean length of pauses	No. of filled Pauses /min	No. of disfluencies /min	No. of stressed words /min	Ratio of words and stressed words	Accuracy	D-value	No of words
Score	.81**	.36	.74**	.88**	-.09	-.62**	-.16	-.13	.92**	-.09	.76**	.57*	.56*
Speech Rate		.32	.76**	.96**	-.17	-.63**	-.23	-.19	.86**	-.02	.66**	.63**	.52*
Articulation rate			.04	.35	-.43	.04	-.45	-.43	.24	.14	.49	.37	.10
Phonation time ratio				.76**	.71	-.89**	.11	.20	.78**	-.01	.30	.31	.57*
Mean length of runs					-.21	-.63	-.22	-.18	.87**	.02	.67**	.65**	.57*
No. of silent pauses/min						-.32	.69**	.73**	.05	.10	-.20	-.35	-.02
Mean length of pauses							-.15	-.40	.-.78**	.01	-.30	-.31	-.57*
No. of unfilled pauses/min								.71**	.-.21	.22	-.46	-.60*	.06
No. of stressed words /min										-.17	.67**	.65**	.47
Ratio of words and stressed words											-.16	-.09	.51*
Accuracy												.70**	.57*
D-value													.34
Number of words													

* indicates $p < 0.05$
** indicates $p < 0.01$

Table 4 Participants' performance on temporal and linguistic measures

Participant	NS1 Score	NS2 Score	NS3 Score	NNS1 Score	NNS2 Score	NNS3 Score	Speech rate	Phon.time ratio	Mean length of runs	Mean length of pauses	No. of stressed words /min	Accuracy	D-value	No of words
1	2	2	3	2	2	1	95.53	31.03	2.73	1621.95	13.43	.0.54	51.36	83
2	3	3	4	2	3	2	164.43	58.75	4.50	843.10	15.84	0.60	33.00	130
3	4	2	3	2	2	2	75.37	35.54	2.70	1267.10	14.62	0.67	41.55	125
4	4	3	4	2	3	2	130.54	66.92	3.55	551.71	17.99	0.19	35.81	191
5	4	4	2	3	2	3	114.08	63.46	3.31	622.51	18.41	0.48	37.97	210
6	3	3	4	4	3	2	99.16	43.49	3.03	1048.84	17.63	0.86	28.76	126
7	4	4	4	4	3	3	142.36	66.29	4.66	650.98	24.43	0.78	48.24	341
8	3	4	4	3	4	3	112.44	45.78	3.68	1094.19	15.81	0.50	32.93	219
9	4	4	4	3	4	3	135.43	67.83	4.47	626.28	17.91	0.46	25.40	191
10	4	4	4	5	4	3	169.46	63.55	6.29	819.24	27.62	0.88	75.78	278
11	5	4	5	5	4	3	157.65	64.46	5.26	547.98	27.89	0.83	37.71	160
12	5	5	4	5	4	4	164.15	63.71	5.40	706.36	29.12	0.83	58.16	124
13	5	4	5	5	5	4	235.48	73.11	7.40	515.58	32.76	0.89	61.79	355
14	5	5	5	5	5	4	205.68	72.31	8.03	610.74	37.61	0.84	61.21	204
15	5	5	5	5	5	5	200.95	66.86	5.93	647.77	49.97	0.97	63.45	197
16	5	5	5	5	5	5	173.86	82.16	6.94	422.07	35.64	0.87	52.80	229