SPEECH RHYTHM PRODUCTION IN THREE GERMAN-ENGLISH BILINGUAL FAMILIES

Nicole Whitworth

Abstract

Rarely, if ever, in studies of the acquisition of more than one phonology has the speech of all the members of a bilingual family been examined within the same experiment. Rather it has been tacitly assumed that the parents' speech complies with the phonetic or phonological characteristics of their respective native languages. For example, the impact of the parents' second language on their native language, and regional and/or idiosyncratic features of the parents' speech have not been taken into account, when evaluating the children's production. However, these possible discrepancies from the standard pronunciation might explain the children's performance, particularly in the non-dominant language. An examination of the speech of parents and children will also provide the opportunity to compare L2 phonology with (developing) bilingual phonology. The experiment reported here compares the speech rhythm of utterances produced by the members of three German-English bilingual families. The children and adults were recorded during a story-telling task. The recordings were then analysed auditorily and acoustically. Rhythmic variability (*Pairwise Variability Index*) was calculated for intervocalic and vocalic intervals of the children's utterances in both languages. The results show that bilingual children are in fact aware of fine-grained rhythmic variability in the linguistic input they receive, and are able to produce corresponding patterns which are, however, not necessarily identical with adult targets.

1 Introduction

Investigations of the acquisition of speech rhythm by first and second language learners are very rare. The main reason for this seems to be the somewhat elusive nature of speech rhythm and how it can be quantified. Recent research by Ramus, Nespor & Mehler (1999), Low, Grabe & Nolan (2001), and Grabe & Low (2002) has been aimed towards the development of an acoustic correlate of speech rhythm which permits the comparison of rhythm in real speech data.

In this study the Pairwise Variability Index (PVI) is used to investigate the production of German and English speech rhythm in three bilingual families. The focus of the study is on the bilingual first language acquisition of speech rhythm by the children of the families. Investigating the speech of their parents within the same study allows a comparison of the production abilities of native speakers, second language learners, and bilinguals who have acquired both languages from birth. A comparison of children's and parents' productions of speech rhythm will, it is suggested, aid in the interpretation of the children's data and also give valuable insights into the different tasks language learners have to face when acquiring a first or second language.

First an overview of previous research into the nature of speech rhythm (Section 2) and how speech rhythm can be measured (Section 3) will be given. Then follows an outline of our knowledge of how speech rhythm is acquired (Section 4). The specific task involved in acquiring English and German speech rhythm is described by presenting a concise contrastive analysis of rhythm in the two languages (Section 5). Finally, the research questions (Section 6)and the experiment set-up and its results are presented and discussed (Sections 7 to 11).

Nelson, D. (ed.) *Leeds Working Papers in Linguistics and Phonetics* 9 (2002), pp. 175-205.

2 Speech rhythm

Rhythm is a regularly recurring sequence of events or movements, which include a beat or stress. There has been a longstanding debate in studies of rhythm whether the beat or the regular recurrence is more important (time vs. accent controversy; for a discussion see Adams (1979, 9 ff.)). However, it seems likely that both are equally important. A third characteristic of rhythm is that it creates the "expectation that the regularity of succession will continue" (Abercrombie 1967, 96).

Rhythm imposes structure on sequences. It has an organising function (Allen & Hawkins 1980). Speech rhythm organises speech into regularly occurring temporal units or events (Fox 2000). In this way the predictability of speech events is increased and thus the intelligibility of utterances (Lehiste 1970). Research has shown that listeners are particularly sensitive to stressed syllables and able to predict when they will occur (Cutler & Foss 1977). This allows recipients to concentrate attention to these events which highlight the semantically significant parts of the utterance and removes the need of constant attention to any speech input. "Rhythmic structure thus produces useful perceptual redundancy in speech by constraining the time when (important) articulatory events may occur" (Allen & Hawkins 1980: 229).

However, the existence of rhythm in speech, that is of isochrony of some sort, has not been an undisputed fact. The debate about whether there is rhythm in speech goes as far back as the second half of the 18th century. In his 1774 paper on the origin of languages Lord Burnet cast doubt on the existence of rhythm, a phenomenon he considers to be something that if existing at all only scholars can perceive (Burnet 1774, 176 ff.). One year later Steele asserted the existence of rhythm in both poetry and prose, likening the succession of heavy and light syllables in speech to the up and down motion of the human foot when walking (Steele 1775, 87 ff.).

A scientific framework for the treatment and classification of speech rhythm was only developed in the last century. Pike (1945, 35) distinguished between two types of isochrony, syllable-timing and stress-timing. Syllable-timed languages are said to have syllables which are of approximately even duration, whereas in stress-timed languages foot durations, the intervals from one stressed syllable to the next, are roughly equal. Abercrombie (1967, 36) explains the existence of rhythm in speech with the breathing mechanism. His theory of chest and stress pulses is based to some degree on Stetson's *Motor Phonetics* (1951). The varying combination of "chest-pulses" and "stress pulses" determines what rhythm a language has. Abercrombie goes one step further than Pike, in declaring that all languages are either stress-timed or syllable-timed. Halliday (1967, 1970, 1985) distinguishes between foot-timing and syllable-timing. He acknowledges that English is not as isochronous in natural speech as it is in counting or rhymes (Halliday 1985, 272). Furthermore, he observes accent-specific variation. British and Australian speech is rhythmically more regular than American or Canadian speech. He concludes that isochrony is phonological and, thus, the phonetically irregular realisation of interstress intervals can be disregarded (Halliday 1967, 12).

However, instrumental studies have further questioned the existence of isochrony. An instrumental phonetic investigation by Classe (1939) analysed the durations between English stressed syllables by means of a kymograph. The results showed that for strict isochrony to occur the syllables of a rhythmic group had to be similar with regard to number as well as phonetic and grammatical structure (Classe 1939, 85). He concluded that true isochrony must be rare in English, "as it may only occur through a complicated system of coincidences" (Classe 1939, 85 ff.), but continues that it "is still a characteristic which always seems to be present and to

make its influence felt; although, frequently, it only remains as an underlying tendency of which some other factor at times almost completely obliterates the effects." (Classe 1939, 90). Later instrumental investigations by Shen & Peterson (1962), Bolinger (1965), O'Connor (1965, 1968), and Uldall (1971) have all provided additional evidence for Classe's (1939) findings. For a review of these studies see Lehiste (1977).

There is, however, reason to re-evaluate some of the conclusions drawn from the results of these instrumental studies. Lehiste (1977, 256) points out that although interstress intervals clearly vary with regard to absolute length, it has to be taken into account that much of the durational variation is below the perceptual threshold. Differences which are below the perceptual threshold, i.e. 10% of the unit for unit durations of 300-500 ms, are perceptually irrelevant. The rhythm of these utterances are, in fact, perceptually isochronous.

Recent research also suggests that there is no one-to-one correspondence between the acoustic signal and linguistic units. This very likely also applies to rhythmicity in speech (Couper-Kuhlen 1993, 14). The discovery of perceptual centres in speech is based on the notion that listeners perceive isochrony in English speech even if there is no evidence for isochrony contained in the acoustic signal (Couper-Kuhlen 1993, 14 ff.; Lehiste 1977, 258 f.). Listeners matching the rhythm of a tone sequence to the rhythm of a sequence of stressed vowels in an English utterance reproduce the tone sequence more regular than the rhythm of the English utterance in reality was (Buxton 1983, 111-112). Listeners tend to overestimate the duration of shorter syllables and underestimate the duration of longer ones (Lehiste 1977, 259). Isochrony in speech rhythm seems to be a psychological reality to speakers as well as listeners. In slips of the tongue speakers of English, a stress-timed language, change the rhythmic make-up of an utterance by deleting or inserting syllables so that syllables are more equally distributed. More commonly, there is evidence that speakers modify segment durations in order to achieve a more isochronous rhythm (Lehiste 1977, 259, f.).

Scott, Isard & Boysson-Bardies (1985) found the same rhythmic tendencies for native speakers of French, a syllable-timed language, and speakers of English. Roach (1982) investigated supposedly stress- and syllable-timed languages. He found that the variability of interstress intervals is no greater in syllable-timed languages than in stress-timed languages. This seems to indicate that this rhythmic phenomenon is a universal property of the human perceptual apparatus, rather than one specific to either stress- or syllable-timed languages (see Section 3). The dichotomy between stress- and syllable-timed languages is therefore also problematic. Furthermore, it has not been possible to assign all languages to either one of the two classes, e.g. Catalan and Polish (Nespor 1990; Ramus, Nespor & Mehler 1999, 269; Grabe & Low to appear).

Dauer (1983, 1987) proposed a different view of speech rhythm class. Like Roach (1982) she concluded that the evidence provided by instrumental studies means that the difference between so-called stress-timed and syllable-timed languages cannot be found in the duration of interstress intervals. Instead the rhythmic make-up of a language is determined by its structural characteristics, i.e. its syllable structure, its stress system, and whether or not it has vowel reduction. Complex syllables are usually found in languages which have traditionally been classed as stress-timed (Abercrombie 1967, 98; Dauer 1983, 55 ff.). Similarly, vowel reduction is found mainly in languages which are considered to be stress-timed (Dauer 1983, 57 f.). The stress systems of stress-timed languages tend to be based on lexical stress which

indicates syllable prominence by changes in length, pitch, loudness, and/or quality, whereas syllable-timed languages tend to have no word level stress at all or rely on pitch alone to mark prominent syllables. A complex set of rules for the realization of stress is thus often a characteristic of stress-timed languages. Dauer's (1983, 1987) account also makes it possible to accommodate rhythmically mixed languages, which have some features that are associated with stress-timing and others that are associated with syllable-timing.

3 Acoustic correlates of rhythm

Ramus, Nespor & Mehler (1999) developed an acoustic correlate of rhythm class based on the segmentation of speech into vowels and consonants. It is derived from the behaviour of newborns who appear to be able to distinguish between different rhythm types without any knowledge of language-specific phonological properties, such as syllable structure and stress system. This means, they argue, that it must be possible to define speech rhythm in phonetic terms alone (Ramus, Nespor & Mehler 1999, 270). Infants' speech perception is characterised by "a succession of vowels of variable durations and intensities, alternating with periods of unanalyzed noise (i.e. consonants)" (Ramus, Nespor & Mehler 1999, 270). They propose that the proportion of vocalic intervals (%V) and the standard deviation of intervocalic interval duration ($\triangle$IV) in a sentence as direct correlates of syllable structure and thus of rhythm class (Ramus, Nespor & Mehler 1999, 274). A low vowel percentage and high standard deviation in intervocalic interval length is associated with stress-timing. The main criticism levelled at this method of measuring speech rhythm is that it does not take account of changes in speaking rate (Grabe & Low to appear). Grabe & Low (2002) point out that the same standard deviation value is produced by a language in which three long vowels are followed by three short vowels and a language in which short and low vowels alternate. Therefore these two languages would be considered rhythmically similar, although they pattern quite differently with regard to vowel duration.

Low & Grabe (1995), Grabe & Low (2002) and Low, Grabe & Nolan (2001) developed an alternative acoustic correlate for speech rhythm class. The *Pairwise Variability Index* (PVI) measures rhythmic variability by calculating the mean difference in the duration of successive intervals (see Section 10.1.2). The PVI is calculated separately for successive vocalic and intervocalic intervals. Vocalic intervals are defined as the stretch of the speech signal which is marked by the presence of a formant structure. Intervocalic intervals stretch from the offset of one vocalic interval to the onset of the next. Grabe and Low (to appear) suggest the application of a normalising procedure for vocalic intervals, since their duration is very much affected by changes in speech rate (Gay 1981). Normalisation for speech rate is not as straightforward for intervocalic intervals. They can consist of different elements, e.g. plosives, fricatives, etc., which are affected differently by speech rate changes. Therefore, a normalisation procedure is not applied to intervocalic intervals.

A high PVI corresponds to a great amount of rhythmic variability which is typical for stress-timed languages. The variability of vocalic intervals will be high in those languages which have vowel reduction. Vowel reduction has generally been held to be a characteristic of stress-timed languages, such as English and German (Dauer 1983). High intervocalic variability is due to complex syllable structures which permit a greater variation in the number of consonants around the syllable nucleus. Complex syllable structure is a characteristic of languages that are traditionally classified as stress-timed (Abercrombie 1967, Dauer 1983).

Although the development of an acoustic correlate for speech rhythm is still in its beginning stages, it seems to be a promising step towards understanding the rhythmic structure of languages. Measures such as the PVI are a useful means of quantifying and thus supplementing auditory impressions of speech rhythm. They facilitate the direct comparison of rhythmic patterns in speech data. However, further research is urgently required. Amongst other things normative data has to be collected and a perceptual basis for the PVI has to be established, i.e. determining the effect of a specific change in the PVI value on a listeners perception of speech rhythm.

4 Acquisition of speech rhythm

Rhythm as a language-specific linguistic feature has to be acquired by language-learning infants. It has been said that it is one of the earliest aspects of speech infants acquire and the most difficult one for adults to modify (Abercrombie 1967, 36). Research in this field has so far been comparatively scarce for monolingual first language acquisition and second language acquisition, and is – to my best knowledge – virtually non-existent for bilingual first language acquisition. The scarcity of investigations reflects the difficulties researchers encounter when studying an elusive prosodic phenomenon like rhythm. The existence of a quantifiable measure of speech rhythm such as the PVI should make it easier for researchers to compare the productions of language learners at various stages and therefore aid the understanding of the acquisition process. Similarly, bilingual data, as is done in the present study, can be compared relatively objectively. The few studies which are available are reviewed below.

4.1 L1 acquisition

Allen & Hawkins (1980) state that children display a bias towards stressed syllables, syllable-timing, and falling (trochaic) accent. The task for children who are acquiring a language with high vocalic variability is thus to learn how to produce reduced syllables. Konopczynski (1995) and Grabe, Post & Watson (1999) found that children learning French (traditionally syllable-timed) acquired the speech rhythm of their language earlier than children learning English (traditionally stress-timed). These findings support Allen & Hawkins' earlier results. Konopczynski (1995) concluded that the acquisition of rhythm is linked to the complexity of rhythmic patterns, i.e. predictability of stress location, naturalness of final lengthening, to which the child is exposed. Grabe, Gut, Post, & Watson (1999) and Grabe, Post & Watson (1999) examined the acquisition of rhythm in English, French and German in the productions of four-year-olds. They showed that the rhythm (i.e. variability of vocalic intervals) of French is acquired earliest followed by German and finally English. Like Konopczynski they deduced that there is likely to be a correlation between rhythmic complexity and age of acquisition. Thus, children appear to acquire speech rhythm moving from a structurally less variable to an increasingly variable rhythmic pattern.

However, it has to be considered that the acquisition of segmental duration and the acquisition of rhythm are implicitly linked (Nespor, Ramus & Mehler 1999, Kent 1976). Segmental durational patterns will have a mutual effect on the rhythmic appearance of a language. The duration of vocalic segments is generally thought to have been acquired by age 3 in stressed and by age 5 in unstressed syllables (Kent 1976). The production of consonant duration becomes adult-like only by age 8 in stressed and by age 10 in unstressed syllables (Kent 1976). It is unlikely that the rhythmic pattern of a language can be fully acquired before the acquisition of segmental duration has become adult-like.

4.2 L2 acquisition

As has already been mentioned above there are only very few studies of L2 rhythm acquisition. One of the most comprehensive is Adams (1979). She studied the acquisition of English speech rhythm by a variety of foreign learners. The auditory and instrumental analysis of her data showed that L2 learners' non-native rhythm is due to a variety of factors, such as insufficient durational difference between unstressed and stressed syllables, i.e. unstressed syllables are not reduced appropriately, missing native linking mechanisms, inappropriate pauses, and misplaced stress. Adams' study thus demonstrate that, at least to some degree, the production of rhythm by second language learners is influenced by their ability to acquire L2 timing patterns, such as are captured by the PVI. These factors are the acquisition of the durational difference between stressed and unstressed vowels, affecting the variability of vocalic interval duration, and the acquisition of native linking mechanisms which involve the loss or insertion of consonant elements, affecting the variability of consonantal interval duration.

4.3 Bilingual first language acquisition

There do not appear to be any studies investigating rhythm acquisition by simultaneous bilinguals.

5 Rhythm in English and German

To get a better understanding of the rhythmic similarities and dissimilarities of the rhythmic structure of the two languages the children and adults are acquiring, this section provides a concise contrastive analysis of rhythm in German and English.

Both British English and German have frequently been described as stress-timed languages (Abercrombie 1967, 97; Buxton 1983, 11; Dauer 1983, 56; Laver 1994 529; Hakkarainen 1995, 151; Kohler 1995, 116). Although they belong to the same rhythm category, their rhythmic structure varies to some degree. The exact rhythmic make-up of a language is determined by the complexity of its syllable structure, the complexity of its stress patterns, and the degree of vowel reduction (Dauer 1983). (See Section 1.) In the following sections the rhythmic structure of English and German is outlined with regard to these aspects.

5.1 Syllable structure

English syllables can have up to three elements in the pre-nuclear syllable margin and up to four elements in the post-nuclear syllable margin, giving (CCC)V(CCCC) (Abercrombie 1967, 75). German syllables can have up to three elements preceding the nucleus and up to five following the syllable nucleus, giving (CCC)V(CCCCC) (Kohler 1995, 176). In a corpus of 8000 syllables Delattre (1965, 42) counted 15 different syllable types for German which ranged from V to CCVCCC, and 14 for English with structures varying from V to CCCVCC. German syllables have on average approximately 2.2 segments per syllable compared to 2.0 in English (Delattre 1965, 42). German has an average of 1.2 consonants per syllable, whereas English has only 1.0 consonants per syllable (Delattre 1965, 41). English thus has a slight tendency to use shorter syllables (Delattre 1965, 42) and has a less variable syllable structure than German.

5.2 Stress patterns

In German and English a syllable is perceived to be prominent when it is marked by pitch movement, higher intensity, and/or longer duration. In German vowel duration

appears to be more significant, whereas pitch seems to be more important in English (Delattre 1965, 33). However, Kaltenbacher (1997, 161) claims that pitch is more important in German and syllable duration and intensity in English. This contradiction might be explained by accent-specific stress patterns.

Stress-placement in words is highly variable in English and somewhat less variable in German (Delattre 1965, 29 ff.). German words and phrases thus adhere to a greater extent to an alternating stress pattern (Kaltenbacher 1997, 159-160).

5.3 Vowel reduction

Delattre (1981, 93) reports that vowel reduction is stronger in English than in German. English unstressed vowels tend to be shorter and more central than German unstressed vowels (Delattre 1981, 92). Typically, in English the reduced form of vowels is the central vowel /ə/, which occurs in the majority of unstressed syllables. The quality of German reduced vowels is more peripheral and thus closer to the quality of the corresponding full vowel (Delattre 1981, 92). In German reduced vowels only occur in word endings and inflectional morphemes (Kaltenbacher 1997, 160).

5.4 Summary: rhythmic make-up of German and English

Syllable structure is more variable in German than in English. Therefore, the German intervocalic PVI is expected to be higher than the English intervocalic PVI, since the variability of non-vocalic intervals is greater. Vowel reduction on the other hand is more marked in English. Consequently the variability of vocalic intervals will be greater in English, resulting in a higher vocalic PVI. Stress-placement is more variable in English. It sounds less regular and successive interstress intervals contain a variable number of unstressed syllables. Hence English will sound less stress-timed than German (Dauer 1983, 58).

Table 1: Rhythmic make-up of German and English based on syllable structure, stress system, and degree of vowel reduction.

	ENGLISH	GERMAN
SYLLABLE	• (CCC)V(CCCC) • short • variable	• (CCC)V(CCCCC) • longer • highly variable
STRESS	• word stress • highly variable	• word stress • variable
VOWEL REDUCTION	• short vowels, central quality • occurs in most unstressed syllables	• longer vowels, more peripheral quality • occurs only in word endings & inflectional morphemes
RHYTHM CLASS	• stress-timed	• stress-timed

The above predictions have been tested in two studies carried out on English and German monolingual adults by Grabe, Gut, Post & Watson (1999) and Grabe & Low (2002). The results are presented in Table 2. Grabe, Gut, Post & Watson (1999) have investigated speech rhythm productions of English, German and French monolingual mother-child pairs. They found significant differences between the normalised vocalic (nVOC) PVI of the German and English mothers. The German mothers have a significantly lower vocalic PVI than the English mothers, which is in

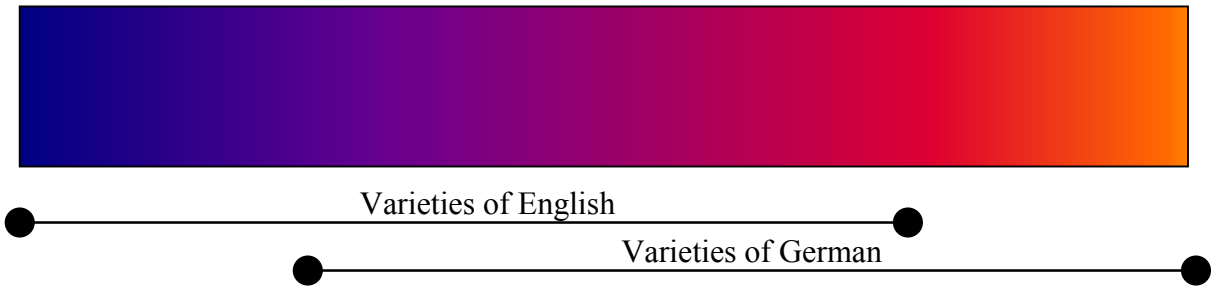
line with the claim that German has a smaller degree of vowel reduction and thus less variability in its vocalic intervals. Raw consonantal (rINT) values were only calculated for the speech of the English mothers, and therefore a comparison is not possible. However, it can be said that the English participants' intervocalic variability is high having an average rINT value of 80.6. Grabe & Low (2002) examined the rhythmic differences in a number of languages based on the speech of a single monolingual subject. In this study the differences between the English and the German productions were not significant. The nVOC values of the German speaker are slightly higher than those of the English speaker, whereas the rINT values are lower than those of the German counterpart. The results of Grabe & Low (2002) do not confirm the predictions made above. Rather they show English to be more variable with regard to consonantal intervals and less variable with regard to vocalic intervals.

Table 2: PVI values for German and English after Grabe, Gut, Post & Watson (1999)[1] and Grabe & Low (2002).

	Grabe et al. (1999)		Grabe & Low (2002)	
	nVOC	rINT	nVOC	rINT
English	68.6	80.63	57.2	64.1
German	49.3	---	59.7	55.3

An explanation of the contradictory results presented above can be found when sociolinguistic factors are considered. The within-language differences are very likely due to regional, sociolectal, and/or gender variations between the speakers who provided the data. The studies have shown that there are significant differences in the rhythmic variability between some speakers but not others, i.e. some varieties of German and English are significantly different with regard to speech rhythm, whereas others are not. This can be represented in form of a graph as below (Figure 1). Unfortunately, since the PVI is a relatively new measure, there is not yet enough normative data available to confirm or contradict this hypothesis. However, it seems to be the most likely explanation.

Figure 1: Overlap of rhythmic patterns in German and English. Rhythm structure is different for the German and English accents at the far ends of the continuum, whereas the varieties which occupy the middle of the continuum are similar.



[1] The values given here are the results of a recalculation of the measurements Grabe, Gut, Post & Watson (1999) made in the study and do not appear as such in the paper.

6 Research questions

- Do bilingual children develop two different rhythmic patterns for their languages, even if these are rhythmically closely related?
- When do bilinguals begin to produce adult-like rhythmic patterns?
- To what extent are their rhythmic productions influenced by their parents' speech, if at all?
- How do the productions of L1 and L2 speakers relate to the productions of bilingual children?

7 Subjects

As subjects three German-English bilingual families from the county of West Yorkshire (England) have been chosen to participate in the study. All the children in the study have been exposed to German and English from birth. All the mothers are native speakers of German and all the fathers native speakers of British English. Both parents are able to speak the other parent's language to some degree (except for [Family C], see section 7.3.1 for details). The families have largely followed a one-language-one person approach, i.e. the mother speaks to the children in German and the father in English.

The dominant language of the children's environment is English, since most of their everyday communication is carried out in that language, e.g. school, friends, etc. Apart from their mother, the children also speak German with their parents' German acquaintances, when they are on holiday in Germany, or telephoning with German relatives. Consequently, their exposure to German is significantly lower than their exposure to English. A more detailed description of the families and their educational and linguistic backgrounds is given below (section 7.1 to 7.3).

In bilingual terminology the children can thus be described as 'simultaneous' and 'co-ordinate' bilinguals, since they have acquired both languages at the same time and from different persons (Lyon 1996, 48; Fantini 1985, 30). According to Romaine's classification the children are examples of Type 1 bilingualism (Romaine 1995, 183ff.).

Table 3: The families. (The parents' L1 language is given in brackets, as is the children's age at the time of the recordings in the format years;months.)

	[Family A]	**Family [B]**	**Family [C]**
Father	Richard (English)	Mike (English)	Martin (English)
Mother	Sabine (German)	Karin (German)	Gabriele (German)
Child 1	Anneliese (7;6)	Rieke (8;10)	Salome (13;2)
Child 2	Leonore (5;0)	Max (6;2)	Reuben (10;10)

7.1 [Family A]

The members of [Family A] are Richard and Sabine and their two children Anneliese and Leonore. The family currently lives in the town of Halifax, West Yorkshire, England.

7.1.1 The father's educational and linguistic background

- Richard is a native speaker of English. He has a perceptible West Yorkshire accent.
- Richard's second language is German. He has studied German for two years in primary school and again in secondary school up to O-level. Later he lived in

Bochum, Germany for approximately two years. An auditory assessment by the author classifies Richard's German as intermediate and non-native.
- Richard has a university education and works in a managerial position.

7.1.2 The mother's educational and linguistic background
- Sabine's native language is German which she speaks with a Northern Standard German accent.
- Sabine started learning English at secondary school when she was approximately 15 years old. She has lived in England for the last twelve years. For the first two years her residence was in Leeds and after that in Halifax. An auditory assessment by the author classifies Sabine's English as fluent and near-native.
- Sabine also studied Latin, French and Spanish.
- Sabine has a university degree and works as a language teacher (German and Spanish) in secondary and adult education.

7.1.3 The children's educational and linguistic background
- *Country of birth and residence*: Leonore and Anneliese were both born and raised in Halifax, West Yorkshire, England.
- *Age*: At the time of the recordings Anneliese was aged 7;6 and Leonore 5;0.
- *Type of bilingualism*: Leonore and Anneliese have been exposed to German and English from birth. The parents have largely followed a one-parent-one-language approach, but of course the children have heard their parents use the respective L2.
- *Language use*: Their normal social environment is English language based. They speak English with their father, English relatives, at school and with their friends. They also use English when they talk amongst themselves. Leonore and Anneliese mainly speak German with their mother and occasionally with their mother's German-speaking friends. They are also exposed to German when they are on holiday in Germany, an event which occurs several times per year. Both children also have the opportunity to watch German children's television. Until a few years ago they also regularly visited a German playgroup which took place approximately once a month.
- *Pronunciation*: The children both have native-like English accents with a noticeable West Yorkshire quality. Their German has a slight foreign accent which does not manifest itself so much in the pronunciation of individual segments but rather in the suprasegmental sphere, i.e. stress patterns, intonation and speech timing.
- *Literacy*: Both children are able to read and write in English. They cannot read and write German. However, Anneliese, the older child, has started lessons.

7.2 [Family B]
[Family B] consists of Mike and Karin, their daughter Rieke and their son Max. They are currently resident in Bradford, West Yorkshire, England.

7.2.1 The father's educational and linguistic background
- Mike is a native speaker of English. He has a slight West Yorkshire accent.
- Mike also speaks German. He started learning the language when he spent three years at a university in East Germany, where he taught information technology.

An auditory assessment by the author classifies Mike's German as beginner to intermediate and non-native.
- Mike has a university degree and he works as an IT consultant.

7.2.2 The mother's educational and linguistic background
- Karin's first language is German. She has a noticeable Brandenburg accent.
- She has been learning English since secondary school. An auditory assessment by the author classifies Karin's English as fluent and non-native.
- Karin has also learned French and Russian.
- Karin has a university degree. She teaches German at a secondary school.

7.2.3 The children's educational and linguistic background
- *Country of birth and residence*: Max and Rieke were both born and raised in Bradford, West Yorkshire, England.
- *Age*: At the time of the recordings Rieke was aged 8;10 and Max 6;2.
- *Type of bilingualism*: Max and Rieke have been exposed to German and English from birth. The parents have largely followed a one-parent-one-language approach.
- *Language use*: Both children live in a primarily English social environment. they speak English with their father, their English relatives, their friends, and at school. They also prefer to speak English when they are talking with each other. Max and Rieke mainly speak German with their mother. They have the opportunity to speak mainly German on their annual holidays in Germany. Until a few years ago they regularly visited a monthly German playgroup.
- *Pronunciation*: Max's and Rieke's English pronunciation is native-like. Rieke's German is native-like. She, like her mother, has a Brandenburg accent, although it is less marked. Max's German has a slight foreign sounding quality which more likely derives from the timing properties of his productions rather than the pronunciation of individual segments.
- *Literacy*: Both children read and write English. Max cannot read or write German. Rieke is also able to read a little German but cannot write it. However, she tends to apply English letter to sound rules when reading German.

7.3 [Family C]
[Family C] are Martin and Gabriele and their two children Salome and Reuben. The family is resident in Bradford, West Yorkshire, England.

7.3.1 The father's educational and linguistic background
- Martin's first language is English. He has a slight West Yorkshire accent.
- Martin does not speak German.
- He has studied French at school.
- Martin has a university degree and is manages his own company.

7.3.2 The mother's educational and linguistic background
- Gabriele is a native speaker of German. She has a Southern Standard German accent.
- She started learning English at secondary school when she was 11 years old. In 1985 she moved to England where she lived in Chester (Cheshire) until 1993 and

then in Bradford. An auditory assessment by the author classifies Gabriele's English as fluent and near-native.
- Gabriele also speaks French.
- Gabriele has a university degree and works as a language teacher in secondary and adult education.

7.3.3 The children's educational and linguistic background

- *Country of birth and residence*: Reuben (10;10) and Salome (13;2) were both born in Chester, Cheshire, England. When the children were aged 2;6 and 6;1 the family moved to Martin's hometown of Bradford, West Yorkshire, England.
- *Age*: At the time of the recordings Salome was aged 13;2 and Reuben 10;10.
- *Type of bilingualism*: Reuben and Salome have been exposed to German and English from birth. The parents have largely followed a one-parent-one-language approach.
- *Language use*: Reuben and Salome are surrounded by a primarily English-speaking environment. They speak English with their father, English relatives, their friends, and at school. They speak German with their mother and their mother's German friends. Approximately once a year they go on holiday to Germany to visit their mother's family. Additionally, Salome has a German friend with whom she has fairly regular telephone conversations.
- *Pronunciation*: Salome sounds native-like in both German and English. Reuben's English is native-like. His German pronunciation has a slight foreign quality which is due suprasegmental rather than segmental features. He has, however, some problems with the pronunciation of the realisation of German /r/ as [ʁ], which he realises as [ɹ]. Superficially this can be explained by interference from English. However, Reuben's German grandmother is reported to have the exact same realisation.
- *Literacy*: Reuben and Salome read and write English. Reuben can read a little German but cannot write it. Salome can read and write German to a certain degree. However, both children tend to apply English letter to sound rules when reading German.

8 Data collection and materials

Tape recordings of the children's speech were made during home visits. The recordings were made on a Tascam DA-P1 DAT machine with lapel microphones. Each subject was recorded on two occasions. In each of these two recording sessions only one language was spoken. The respective monolingual parent was asked to be present, although this was not always possible. These steps were taken in order to control for language mode, that is, it has been attempted to activate a monolingual German mode in the German recording session, and a monolingual English mode in the English recording session (Grosjean 1998). The data were elicited by means of a story-telling task, in which the children were asked to tell a story which was depicted by a series of black-and-white pictures. The story used for the English recordings was Mercer Mayer's *One Frog too Many* and for German *Frog Where are You?* by the same author[2].

[2] The titles of the frog stories are both English, since they are published in the United States. However, since they do not contain any text, it was possible to use them in the experiment as outlined above.

This approach follows the methodology of earlier studies by Grabe, Gut, Post & Watson (1999) and Grabe, Post & Watson (1999). In these studies the data was elicited by means of a proprietary game from the Early Learning Centre ('Treasure Chest of Tales'). The choice of the frog story books in this study was motivated by the desire to use different stories, which are nevertheless of a similar type, for each language in order to separate the linguistic background of the two taping situations further. Moreover, the subject will be guided by the pictures and different subjects will at least produce similar utterances. A third consideration is that the frog story books are a widely-used tool in the field of language acquisition. Their use thus increases the chance that the recordings might be useful to other researchers in the field.

Other investigations of speech rhythm that involve the PVI have relied on read material, which can be controlled better with regard to linguistic variables (e.g. Grabe & Low to appear, Low, Grabe & Nolan 2001). This was not practicable in this study. First, although most of the children had acquired some reading skills in English, especially the younger ones were not sufficiently fluent. Secondly, only a minority of the children and L2 speakers were able to read German and again the level of proficiency was low. Consequently, the use of read material would have misrepresented the children's and adults' production skills. Although connected speech carries with it many problems with regard to segmentation and comparability (see above), it was felt that it would give a more accurate picture of the subjects' speech.

The use of spontaneous speech, however, brings with it a number of problems. Some of these problems relate to the segmentation of the utterances and others to the comparability of the data. Unprepared speech contains a larger number of errors of performance such as hesitations and false starts (Cruttenden 1997, 29; Grabe 1998, 41). It is also much more difficult to segment its constituents. For example, intonation phrase boundaries are more difficult to identify (Cruttenden 1997, 29; Grabe 1998, 41). Moreover, issues of comparability of the two languages arise, since the material cannot be controlled as to the number and complexity of vocalic and intervocalic intervals, size of utterances and intonation phrases, speech rate etc. For a discussion of some of the issues see Grabe (1998, 55).

9 Data analysis

For each subject 30 intonation phrases (IPs), i.e. 15 per language, containing at least four syllables, were extracted from the recorded story-telling material. Intonation phrases which contain any internal pauses or other errors of performance have been excluded from the analysis. The only exceptions of this rule are English IP3, and IP9, and German IP3, IP7, IP9, and IP14 of Leonore, a member of [Family A]. Aged 5;0, she is the youngest participant in the study. Due to the relative immaturity of her speech she did not produce any alternative 'regular' intonation phrases. The errors of performance observed in her speech productions are IP-internal hesitations and in-breaths. See section 9.1.

The phrases were all produced at a relatively constant speaking rate (judged auditorily by the author). To ascertain that the speaking rate was relatively stable and as a measure for speech maturity, mean interval duration of vocalic intervals (VOC MID) and intervocalic intervals (INT MID) were calculated for all utterances. See section 10.1.

As an acoustic correlate for rhythm, the variability of successive vocalic and intervocalic intervals was analysed. This was done by calculating the *Pairwise*

Variability Index (PVI). First the lengths of vocalic and intervocalic intervals were measured on wideband spectrograms generated on a Sensimetrics Speech Station 2. The measurements were then entered into equations for either the *Raw Pairwise Variability Index* (rPVI) or the *Normalised Pairwise Variability Index* (nPVI). See Section 10.1.2.

The rhythmic analysis of the subjects' speech requires the segmentation of the collected linguistic material first into intonation phrases and then into vocalic and intervocalic intervals. Furthermore, the final syllable of the intonation phrase is disregarded in the calculation of the PVI. This is done to eliminate the effects of pre-boundary lengthening, which is a separate acquisition issue that is not discussed here. The sections below outline the criteria applied in the segmentation task. Criteria to determine the boundaries of intonation phrases (section 9.1), vocalic intervals (section 9.2), intervocalic intervals (section 9.2), and of the final syllable of the intonation phrase (section 9.3) are discussed below.

9.1 Segmentation criteria: intonation phrase (IP)

Although the intonation phrase is in some form or other present in most intonational frameworks, it cannot always be unambiguously defined, particularly in spoken conversation (Cruttenden 1997, 29; Grabe 1998, 41). Cues that demarcate the boundaries of an intonation phrase can be phonetic in nature, i.e. external (Cruttenden 1997, 29), or intonational, i.e. internal (Cruttenden, 1997, 29).

External criteria which may indicate an IP boundary are pause, anacrusis, pre-boundary syllable lengthening, and pitch movement on an unaccented syllable.

- *Pauses* can be silent or filled. They regularly occur at a number of places within an utterance and accordingly are divided into three types. Type 1 marks important syntactic boundaries, type 2 occurs before words which carry a high informational content, and type 3 after the initial word of an intonation phrase where pauses are generally due to errors of performance (Cruttenden 1979, 30-32). Pauses of type 2 and 3 tend to be found in IP-internal position, and thus do not tend to demarcate IP boundaries (Cruttenden 1997, 32). A more detailed treatment of speech pauses can be found in e.g. Butcher (1981), Duez (1982).
- *Anacrusis* is a phenomenon which can be observed at the beginning of intonation phrases, where unstressed syllables which precede the first stressed syllable are temporally compressed, i.e. they are uttered more quickly (Cruttenden 1997, 32).
- The term *pre-boundary lengthening* or final-syllable lengthening is used to refer to the increased length of the final syllable before a boundary. The actual degree of lengthening varies for different languages (Cruttenden 1997, 33). For a more detailed discussion of this phenomenon see section 9.3.

The above criteria do not always permit an unambiguous assignation of intonation phrase boundaries. Pauses, as has already been outlined, can also occur IP-internally. The same applies for anacrusis and final-syllable lengthening, which may also indicate errors of performance, i.e. hesitation (Cruttenden 1997, 34). Pitch movement on a syllable which is normally unstressed and can thus not be accented is a reasonably reliable marker for an IP boundary (Cruttenden 1997, 34).

The internal structure of intonation phrases can also help to delimit the boundaries of intonation groups, since every IP has to consist of a minimal number of elements. (Cruttenden 1997, 34). Each intonation phrase has to consist of at least one stressed syllable and must contain at least one pitch movement from or to an accented

syllable (Cruttenden 1997, 35). However, there are a number of problems. First, as Cruttenden (1997, 29) remarks, it is a circular way of determining intonation phrase boundaries, since minimal intonation patterns are defined by their occurrence in intonation phrases on the one hand, and intonation phrases are defined by having a minimal internal intonation structure on the other. Second, there are a number of problematic pitch patterns in English and German which prevent an unambiguous analysis into intonation phrases, e.g. intonational sandhi (Cruttenden 1997, 35 ff.). For a more detailed account of English and German intonation see Cruttenden (1997) and Grabe (1998).

In summary, the following criteria will be used to dissect the subjects' utterances into intonation phrases. They can be divided into features which signal the presence of a boundary and features which are mandatory components of an IP.

- Indicating an IP boundary:
  1. pause;
  2. anacrusis;
  3. final-syllable lengthening;
  4. pitch movement on unaccented syllable.
- Core constituents of IPs:
  5. contains at least one stressed syllable;
  6. contains at least one pitch movement from/to an accented syllable.

9.2 Segmentation criteria: vocalic and intervocalic intervals

The task of dividing the continuous speech signal into segments is a notoriously difficult one (e.g. Peterson & Lehiste 1960, 694; O'Connor 1973, 93 f.; Laver 1994, 101; Fox 2000, 13). This, of course, makes measuring segment length equally difficult, since it is problematic to decide where a segment begins and where it ends. The definition of clearly defined and generally acceptable segmentation criteria is thus necessary to ensure the comparability of the data, not only within this study but also to the data of other studies. The segmentation procedure in this study will therefore follow the criteria outlined by Peterson & Lehiste (1960 694 ff.) and Grabe & Low (to appear, 13 ff.).

Vocalic intervals are defined as the portion of the speech signal between the onset and the offset of a vowel, in the phonemic sense of syllable nucleus. They may stretch over more than a single syllable and even across word-boundaries, since the PVI is based on acoustic rather than phonological principles. For example, Sabine – one of the German mothers – produces the sequence shown in example (1).

(1) [Family A], Sabine, German IP 01
[zeːn.vɪɐ.aɪ.n.klai.n.jʊŋ.n]
sehn wir einen kleinen Jungen
'we see a little boy'

This sequence is split into the following vocalic intervals: [eː], [ɪɐ.aɪ.n ‿], [ai.n̩], and [ʊ]. The fifth vocalic interval [n̩] constitutes the final syllable of the utterance and is not included in the PVI calculation (see Section 9.3). In this example, the second vocalic interval stretches over three syllables and the third over two syllables.

On the spectrogram vocalic intervals can be distinguished by the existence of a formant structure. The duration of vocalic intervals was measured from the onset of

the vowel to its offset. Devoiced vowels which do occur in both English and German do not exhibit a formant pattern and are not included in the vocalic portions (Grabe & Low to appear, 15). Liquids, glides, and nasals, which also possess a formant structure, were excluded from the vocalic section, where they could be identified clearly by spectral changes in the acoustic signal.

Intervocalic intervals are defined as the portion of the speech signal between vowel offset and vowel onset, a stretch of speech normally characterised by the absence of vowel formants. Liquids, glides, and nasals, which also have formants, are included in the intervocalic portion, where they could be clearly distinguished from the vocalic portion through spectral changes. An intervocalic interval may contain one or more consonants and may stretch across syllable and word boundaries. The utterance in Example (1) is divided into five intervocalic intervals: [z], [n.v], [kl], [j], and [ŋ]. The second of these intervals stretches across a syllable-boundary, indicated by [.] which is in this case also a word-boundary. The third interval consists of two segments which are part of a syllable onset.

9.3 Segmentation criteria: final syllable

The final syllable of each of the intonation phrases has been disregarded in the calculation of the PVI, since the acquisition of pre-boundary lengthening might be independent from the acquisition of rhythm as such. Severing the final syllable from the rest of an IP raises issues regarding the definition of the syllable and the location of syllable boundaries, which are by no means trivial (e.g. Laver 1994, 113 ff.).

In this study the identification of the left boundary of the final syllable has been based on articulatory principles. In many of the instances the assignation of the boundary to the final syllable was relatively straightforward since it coincided with a word-boundary.

(2) [Family C], Martin, English IP 02
[ɪz.hɪz]
(*which I presume*) *is his*

In other cases the phonotactic rules of the language clearly indicated the syllabic affinity of a segment.

(3) [Family B], Max, English IP 11
[rʊg.bɪ]
(we play at) rugby

In those cases where the affinity of a consonant to either the penultimate or the ultimate syllable was in doubt (e.g. so-called ambisyllabic consonants), the consonant was considered to be part of the syllable which carried the stress (Krakow 1999).

(4) [Family A], Sabine, English IP 07
[sʊd.n]
(all of a) sudden

10 Results

The results are presented below. A total of 4120 intervals have been measured, i.e. 1033 vocalic and 1065 intervocalic English intervals, and 985 vocalic and 1037

intervocalic German intervals. The data were then analysed to provide acoustic correlates of speech maturity and speech rhythm. As a correlate for speech maturity mean interval duration (MID) was used and for speech rhythm the Pairwise Variability Index (PVI).

10.1 Mean Interval Duration (MID)

The mean interval duration of vocalic intervals (VOC MID) and the mean interval duration of intervocalic intervals (INT MID) and their respective standard deviations (sd.) were calculated as a measure of speech maturity. Language acquisition studies repeatedly show that segment duration tends to be greater in developing speech than in mature speech and decreases with increasing age (e.g. Tingley & Allen 1975; Kent 1976). Simultaneously, reliability of the production of segment durations increases (Tingley & Allen 1975).

The development of vowel duration and its consistency is thought to stabilise to an adult-like state between age 3 and 5 first for stressed and then for unstressed vowels (Allen & Hawkins 1980, 237). In contrast consonant duration and consistency only stabilises later at approximately age 10 (Allen & Hawkins 1980, 237). Since the development of vocalic and intervocalic interval duration is directly linked to the development of segment duration it can be expected to come to an end at age 5 for vocalic intervals and at age 10 for non-vocalic intervals.

Correspondingly, VOC MID is expected to be relatively high before age 5 and INT MID before age 10, as is their respective variability (standard deviation of VOC MID and INT MID). A stabilisation of L1 speaker values of both VOC MID and INT MID as well as their standard deviations is predicted for the productions of children above the age of 5 and 10, respectively.

For the purposes of this study, a low mean interval duration is interpreted as an indication of speech maturity. Correspondingly, a low standard deviation of MID is taken as a sign of consistency in duration production and is thus also interpreted as an indication of relative speech maturity.

10.1.1 English MID

The English mean interval durations and corresponding standard deviations have been displayed in Figure 2 and Figure 3 below. The values of the adult females and adult males have been averaged. The children's results are displayed individually.

- Vocalic intervals:
  - English vocalic MID and standard variation are adult-like for the four oldest children who range in age from 13;2 to 7;6.
  - The two youngest children, Leonore (5;0) and Max (6;2) have a higher mean interval duration and standard deviation. However, Leonore's values are slightly lower than Max's, although she is more than a year younger.
  - There are no differences between the performance of the adult L1 and L2 speakers.
- Intervocalic intervals:
  - The intervocalic MIDs of the subjects' English productions decrease with increasing age, i.e. the youngest child has the highest intervocalic mean interval durations.
  - The standard deviations of the MIDs follow the same pattern of with only one exception. Max (6;2) has a low standard deviation which is only slightly higher than those of the adults and the older children.

- Adult L1 and L2 intervocalic intervals have approximately the same mean duration. However, the standard deviation of the L2 speakers is slightly higher.

Figure 2: English vocalic MIDs and standard deviations (in ms).

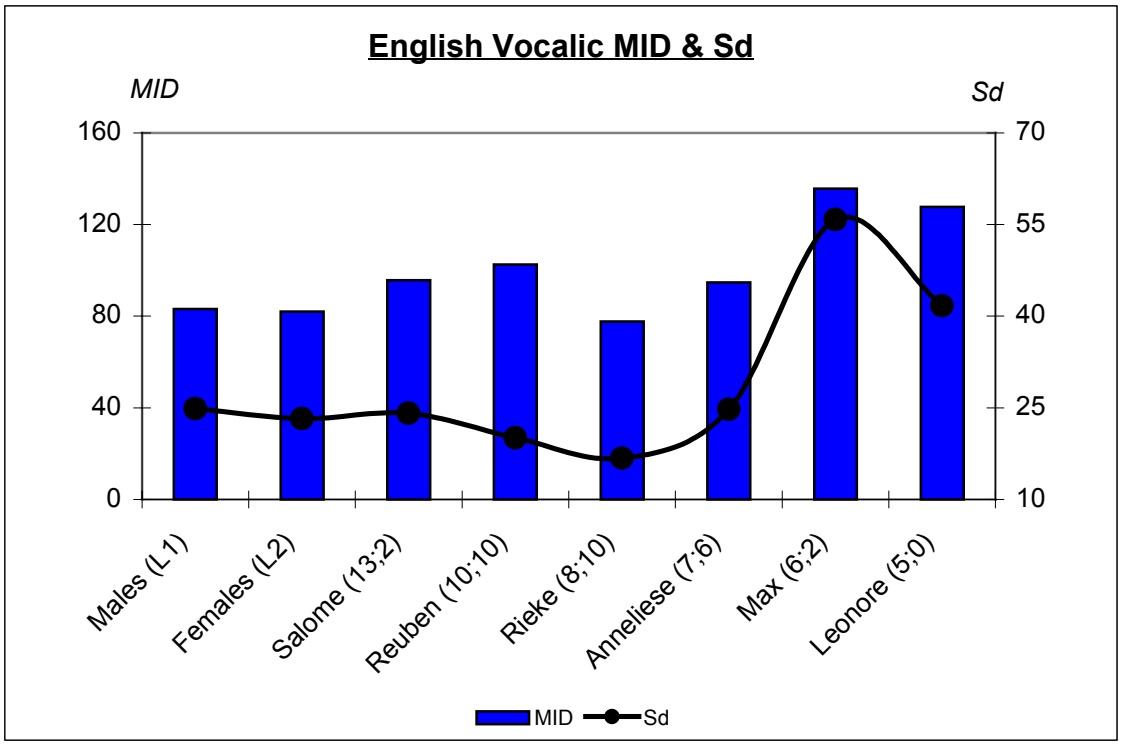


Figure 3: English vocalic MIDs and standard deviations (in ms).

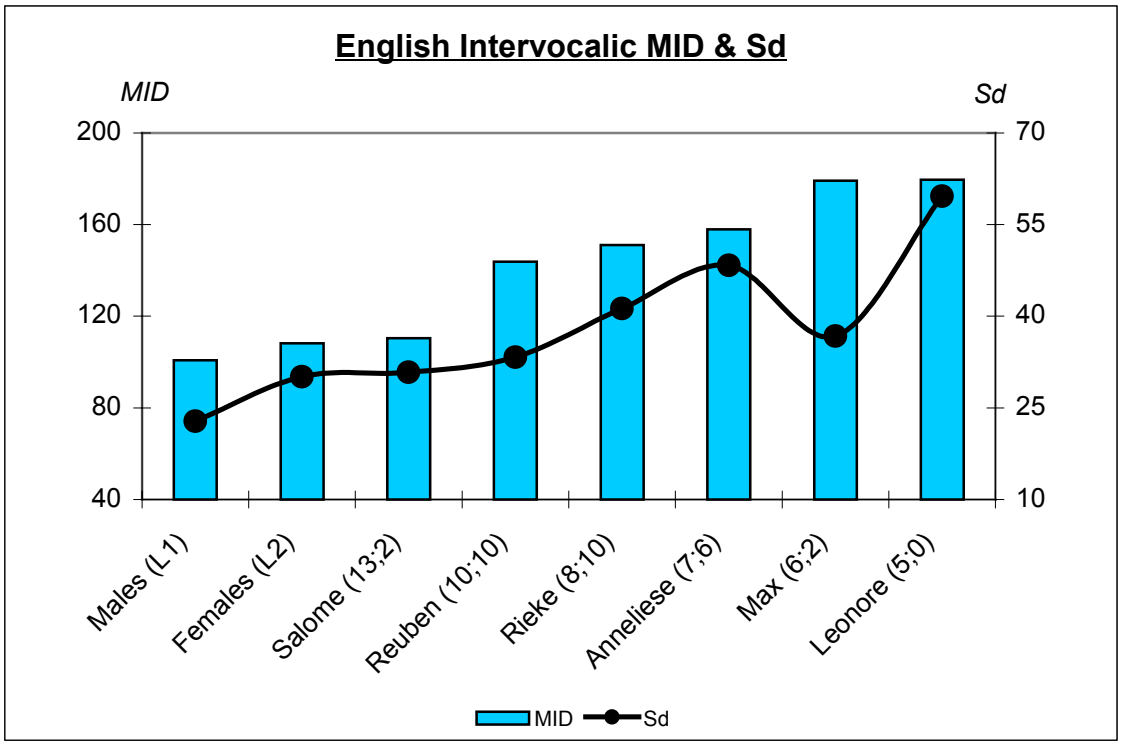
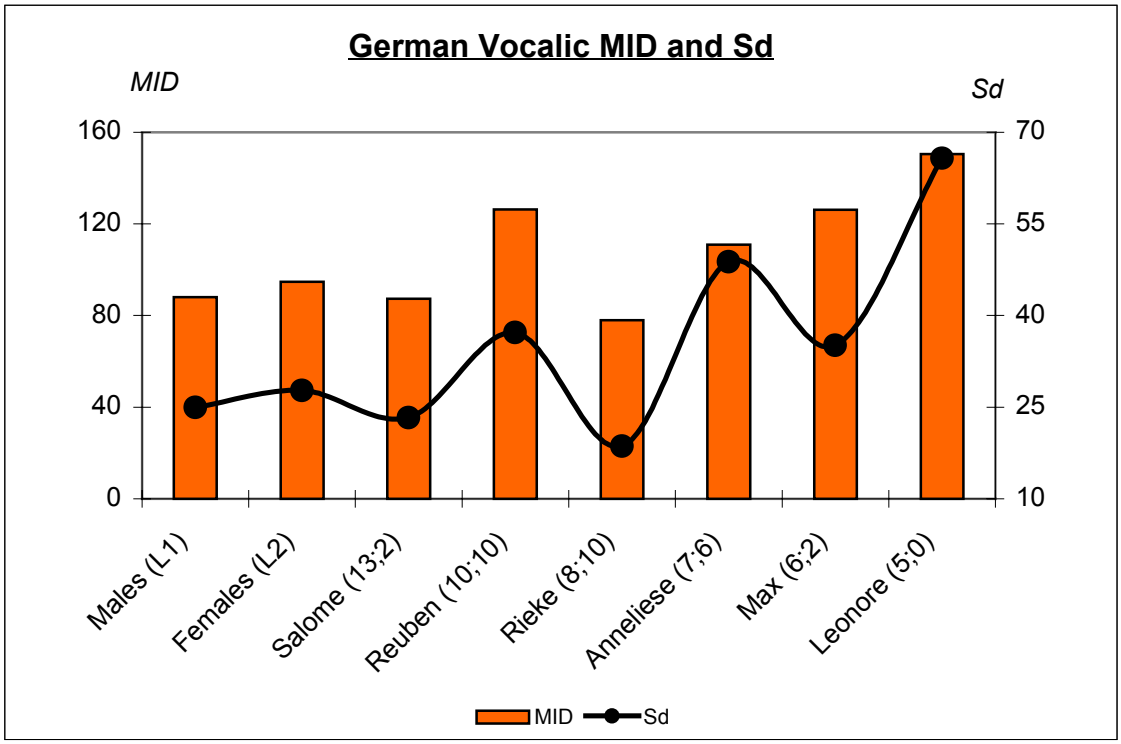


Figure 4: German vocalic MIDs and standard deviations (in ms).



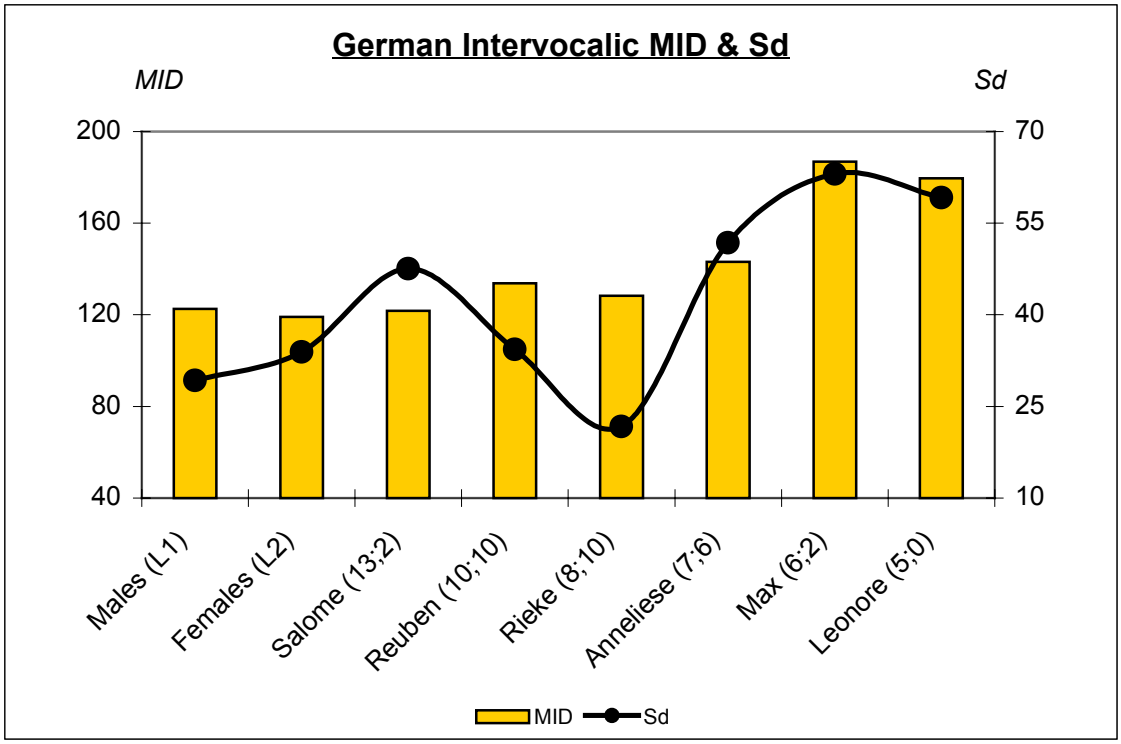


Figure 5: German intervocalic MIDs and standard deviations (in ms).

10.1.2 German MID

The German mean interval durations and corresponding standard deviations are shown in Figure 4 and Figure 5 above. The values of the adult females and the adult males have been averaged. The children's values have been displayed individually.

- Vocalic intervals:
  - The German vocalic MIDs of the two oldest girls, Salome (13;2) and Rieke (8;10) are adult-like.
  - Anneliese (7;6), Max (6;2), and Leonore (5;0) have increasingly high mean interval durations.
  - Reuben (10;10) seems to have unusually high MIDs for his age, since his MID is similar to that of six-year-old Max.
  - The standard deviations of the vocalic MID results follow a very similar pattern. However, Max's standard deviation is lower than that of Anneliese who is approximately a year older. Reuben's values reflect those of Max.
- Intervocalic intervals:
  - German intervocalic mean interval duration appears to be adult-like for the four oldest children: Salome (13;2), Reuben (10;10, Rieke (8;10), and Anneliese (7;6).
  - The two youngest children, Leonore (5;0) and Max (6;2), produced higher MIDs than the adults and older children.
  - The standard deviations of intervocalic MIDs are adult-like for the four oldest children. However, Salome's standard deviation is noticeably high.

10.2 Pairwise Variability Index (PVI)

The PVI is a unit which expresses the degree of rhythmic variability contained in the acoustic speech by determining the mean difference in the length of successive acoustic-phonetic intervals. The greater the mean difference, the higher is the rhythmic variability of the intervals. Thus, a high PVI value corresponds to high variability in the length of successive intervals, which is characteristic of stress-timed languages (see Section 3).

The *Normalised PVI* was used to calculate the rhythmic variability of vocalic intervals: nVOC. Speech rate normalisation is required for vocalic intervals, since vocalic segments are lengthened when the speech rate is decreased and shortened when the speech rate is increased (Gay 1981). Grabe & Low (2002) have shown that the vocalic PVI is linked to the mean length of vocalic intervals. This measure has been used as a correlate for speech rate (see also Section 10.1).

Figure 6: Equation used to calculate the Normalised Pairwise Variability Index for vocalic intervals *(nVOC). Adapted from Grabe & Low (2002).*

$$\mathrm{nVOC} = 100\,\mathrm{x}\left[\sum_{k=1}^{m-1}\left|\frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2}\right| /(m-1)\right]$$

where m = number of intervals
d = duration of the k^{th} interval

The *Raw PVI* was used to calculate the rhythmic variability of intervocalic intervals: rINT. Changes in speech rate do not affect non-vocalic segments to the same degree as vocalic segments (Gay 1981). Moreover, intervocalic intervals can contain several segments which may be affected differently by a faster or slower speech rate. A process of normalising intervocalic intervals for speech rate effects has not yet been developed, due to the complex nature of speech rate effects on non-vocalic segments (Grabe & Low 2002).

Figure 7: Equation used to calculate the Raw Pairwise Variability Index for intervocalic intervals (rINT). Adapted from Grabe & Low (2002).

$$\text{rINT} = \left[\sum_{k=1}^{m-1} \left| d_k - d_{k+1} \right| / (m-1) \right]$$

where m = number of intervals
d = duration of the k^{th} interval

10.2.1 [Family A]

The mean PVI results of [Family A] are shown in Figure 8 below. Richard, the father, is a native English speaker. Sabine, the mother, is a native German speaker. The bilingual children are 7-year-old Anneliese and 5-year-old Leonore.

- English nVOC:
  - Anneliese's English nVOC value is slightly higher than Leonore's nVOC. However, her values do not differ significantly from those of her younger sister Leonore.
  - Sabine's L2 value is lower than her husband's L1 value. Her L2 and Richard's L1 English vocalic PVIs do not differ significantly.
  - The children's English nVOCs are lower than the adults' results. Richard's L1 vocalic PVI is significantly different from his daughter Leonore's English nVOC ($p<0.02$), but they do not from Anneliese's English nVOC. Sabine's L2 English nVOC is not significantly different from those of her daughters.
- English rINT:
  - The children have the same English rINT values.
  - Sabine's L2 value is marginally lower than Richard's L1 result. The results do not differ significantly.
  - The children's values are higher than the adults' values. Richard's L1 result is significantly different from both his daughters' rINT PVIs ($p<0.03$). Sabine's L2 rINT is not significantly different from her daughters' intervocalic PVIs.
- German nVOC:
  - Anneliese's German nVOC PVI is slightly lower than Leonore's value, but not significantly.
  - Sabine's L1 German nVOC is higher than Richard's L2 German nVOC. The difference is not significant.
  - The children's German nVOCs are lower than their parents' L2 German nVOCs. Sabine's L1 German nVOCs differ significantly from those of Anneliese ($p<0.04$). Sabine's nVOCs do not differ significantly from

Leonore's. Richard's L2 German nVOCs do not differ significantly from those of either child.

- German rINT:
  - Anneliese's German rINT values are insignificantly lower than Leonore's German rINTs.
  - Sabine's L1 German rINTs are higher than Richard's L2 German rINT, but not significantly.
  - The children's German rINTs are higher than those of their parents. The difference between Sabine's L1 German rINT and her daughter's German rINTs is not significant. The difference between Richard's L2 German rINTs and his daughters' German rINTs is significant ($p<0.03$).
- English vs. German PVI:
  - Richard's L1 English nVOCs are virtually the same as Sabine's L1 German nVOCs. The difference is therefore not significant.
  - Richard's L1 English rINTs are significantly lower than Sabine's L1 German rINTs ($p<0.04$).
  - None of the members of [Family A] realises significantly different rhythmic patterns (nVOC or rINT) in their German and English speech.

10.2.2 [Family B]

The mean PVI results of [Family B] are displayed in Figure 9 below. Mike, the father, is a native speaker of English. Karin, the mother, is a native speaker of German. The children are 8-year-old Rieke and 6-year-old Max.

- English nVOC:
  - Rieke's English nVOCs are significantly lower than Max's English nVOCs ($p<0.04$).
  - Karin's L2 English nVOCs are lower than Mike's L1 English nVOCs. However, the values do not differ significantly.
  - Rieke's English nVOC is lower than her parents' nVOCs. However, the difference between her and her mother's L2 values is not significant, whereas the difference between her and her father's L1 values is ($p<0.04$). Max's English nVOC is higher than those of his parents, but his results do not differ significantly from either his father's L1 or his mother's L2 results.
- English rINT:
  - Rieke's English rINT is significantly lower than Max's English rINT ($p<0.04$).
  - Karin's L2 English rINTs are higher than Mike's L1 values. The difference is not significant.
  - Both children have a higher English rINT than their parents. Mike's L1 English rINT is significantly lower than the rINTs of both children ($p<0.02$). The difference between Karin's L2 English rINT and her daughter's result is not significant. The difference between Karin's and her son's English rINT is significant ($p<0.01$).
- German nVOC:
  - Rieke's German nVOC is lower than Max's, but the difference is not significant.
  - Karin's L1 German nVOC significantly lower than her husband's ($p<0.03$).
  - Rieke's German nVOC is significantly lower than her mother's L1 German nVOC ($p<0.006$). Rieke's German nVOC is also lower than her father's L2

German nVOC, but the difference is not significant. Max's German nVOC is higher than his parents', however the differences are not significant.

- German rINT:
  - Rieke's German rINT is significantly lower than Max's German rINT ($p<0.01$).
  - Karin's L1 German rINT is lower than Mike's. The two results do not differ significantly.
  - Rieke's German rINT is higher than her mother's L1 German rINT and lower than her father's L2 German rINT. Neither difference is significant. Max's German rINT is significantly higher than his mother's L1 and his father's L2 German rINT ($p<0.02$).
- English vs. German PVI:
  - Mike's L1 English nVOCs are slightly lower than Karin's L1 German nVOCs. The difference is not significant.
  - Mike's L1 English rINTs are marginally lower than Karin's L1 German rINTs. The difference between the values is not significant.
  - None of the members of [Family B] realises significantly different rhythmic patterns (nVOC or rINT) in their German and English speech.

10.2.3 [Family C]

The PVI results of [Family C] are presented in Figure 10. Martin, the father, is a native speaker of English. Gabriele, the mother, is a native speaker of German. The children are 13-year-old Salome and 10-year-old Reuben.

- English nVOC:
  - Salome's English nVOC is lower than Reuben's, but the difference is not significant.
  - Gabriele's L2 nVOC value is higher than her husband's L1 English values. However, the two values do not differ significantly.
  - Salome's and Reuben's English nVOCs are lower than their mother's L2 English nVOC and higher than their father's L1 English nVOC. None of the differences are significant.
- English rINT:
  - Salome's English rINT is insignificantly lower than Reuben's.
  - Gabriele's L2 English rINT is significantly higher than Martin's L1 English rINT ($p<0.02$).
  - Salome's English rINT is lower than her mother's L2 rINT and higher than her father's L1 rINT. Reuben's English rINT is higher than his father's L1 rINT and equal to his mother's L2 rINT. None of the differences are significant.
- German nVOC:
  - Salome's German nVOC is lower than Reuben's. The difference is not significant.
  - (Since Martin does not speak German, a comparison with Gabriele's German nVOC was not possible.)
  - Salome's German nVOC is insignificantly lower and Reuben's German nVOC is insignificantly higher than Gabriele's L1 German nVOC.
- German rINT:
  - Salome's German rINT is lower than Reuben's German rINT. The two values do not differ significantly.
  - (No comparison possible.)

- Salome's German rINT is slightly lower and Reuben's German rINT is higher than Gabriele's L1 German rINT. The differences are not significant.

- English vs. German PVI:
  - Martin's L1 English nVOCs are higher than Gabriele's L1 German nVOCs. The difference is not significant.
  - Martin's L1 English rINTs are marginally lower than Gabriele's L1 German rINTs. The difference between the values is not significant.
  - Both Gabriele and Salome realise significantly different nVOC patterns in their German and English speech, whereas Reuben's German and English nVOC are not significantly different ($p<0.06$). None of the members of [Family C] realises significantly different rINT patterns in the two languages.

Figure 8: Mean PVI of vocalic (nVOC) and intervocalic (rINT) intervals as produced by [Family A] in English and German.

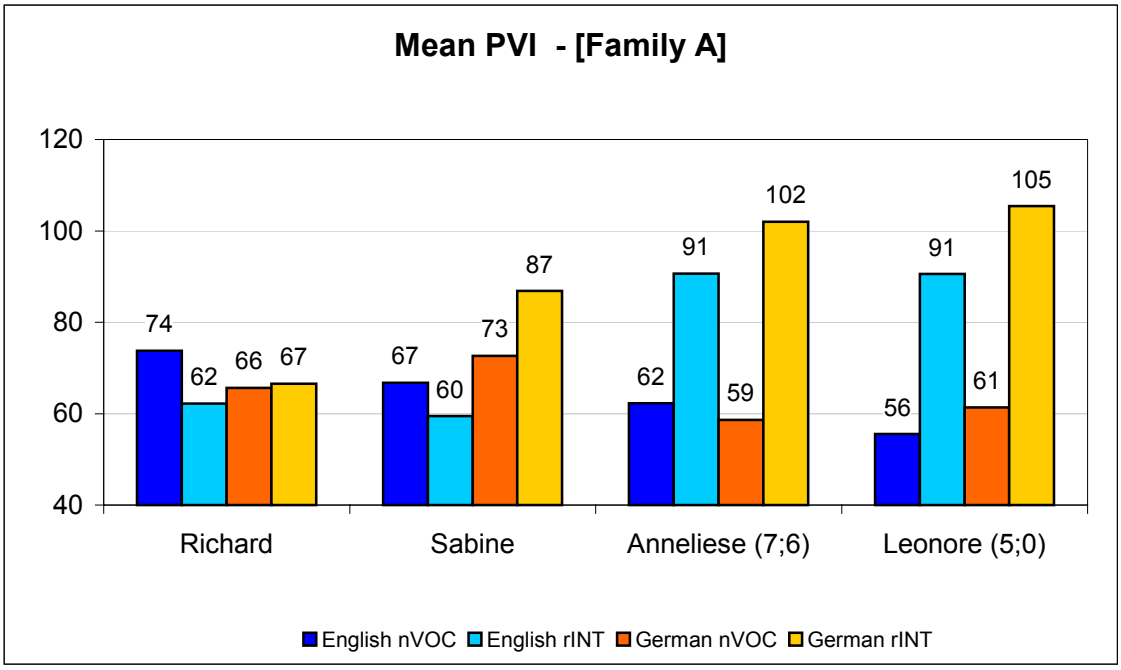


Figure 9: Mean PVI of vocalic (nVOC) and intervocalic (rINT) intervals as produced by [Family B] in English and German.

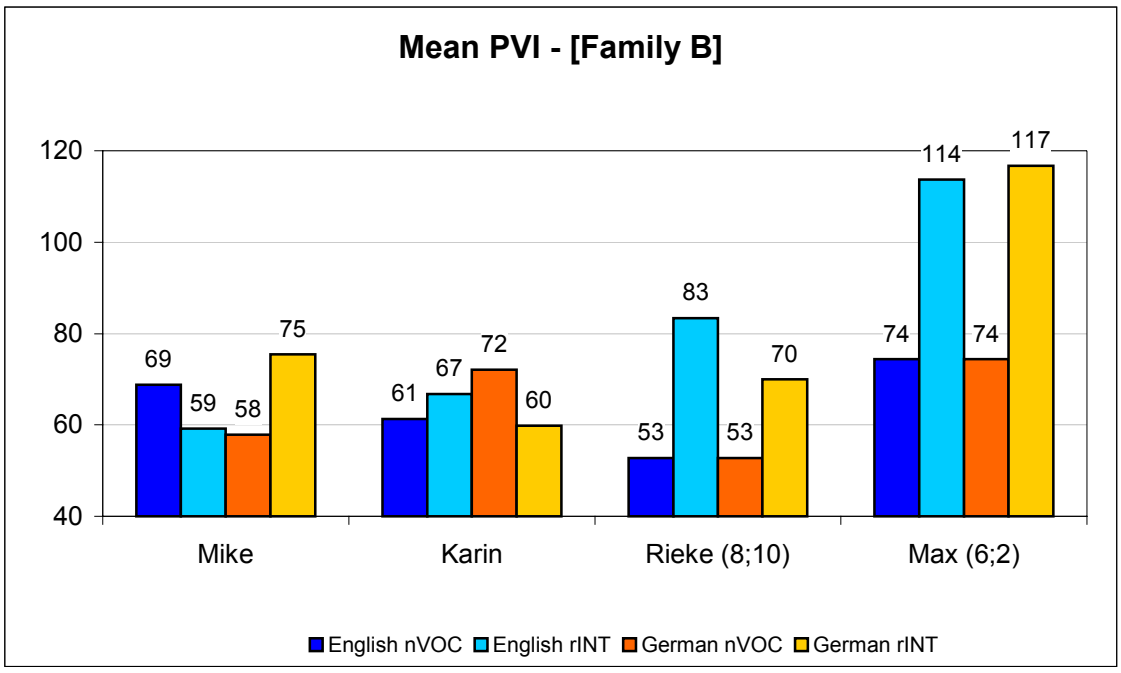
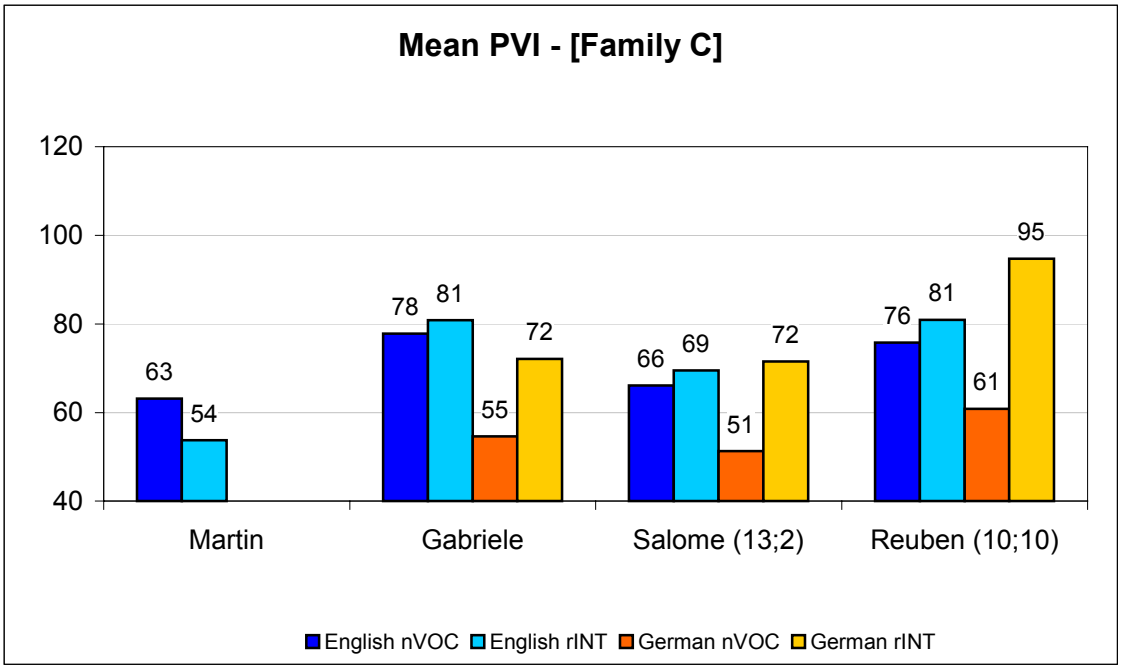


Figure 10: Mean PVI of vocalic (nVOC) and intervocalic (rINT) intervals as produced by [Family C] in English and German.



11 Summary and discussion

11.1 Mean Interval Duration (MID)

The MID results have shown that as expected the duration of intervals decreases with increasing speaker age. At the same time the variability (sd.) with which these durations are produced increases with increasing age. There are a few exceptions to this general trend.

Adult L1 and L2 values do not differ with regard to MID. This is not surprising, since the production of interval durations seems to be – at least to some degree – related to the maturity of the speech motor processes (Kent 1976). Adults can thus be expected to produce comparatively short durations very reliably whether they are speaking a first or a second language.

11.2 Pairwise Variability Index (PVI)

An analysis of the speech data gathered from *[Family A]* produced the following results:

- Anneliese's (7;6) and Leonore's (5;0) nVOCs in both languages are the same or lower than the L1 and L2 patterns of the two adults, whereas the rINT values are higher.
- The children's productions are very similar with regard to rhythmic patterns.
- With the exception of German rINT values the children's productions tend to be significantly different from L1 speech and more like L2 speech.
- Sabine's L2 rhythmic pattern is quite closely matched to her husband's L1 pattern. Richard, however, realises a different pattern in his L2 German speech than his wife in her L1 German.
- The children do not realise different rhythmic patterns for their two languages, and neither do their parents. Furthermore, a comparison of the parents' L1 productions reveals that the rhythmic patterns of Sabine's L1 German and Richard's L1 English are not significantly different.

An analysis of the speech data gathered from *[Family B]* produced the following results:

- Rieke's nVOCs are lower and her rINTs higher than those of her parents. Max's nVOCs and rINTs are higher than those of the adults.
- The PVI values of the two children are quite different. However, when looking at the overall pattern, there still appears to be some similarity, e.g. both have the exactly same nVOC index for both languages (Max's nVOC for both German and English is 74ms and Rieke's 53ms.)
- Although for [Family B] the results are not quite as clear-cut, there also tends to be a trend for the children's rhythmic patterns to be closer to L2 than to L1 values, at least for Rieke. Max's overall very high values result in either making them significantly different from both L2 and L1 speech or not.
- Karin's L2 English PVIs are not significantly different from her husband's L1 values. Whereas Mike's L2 German nVOC differs significantly from Karin's L1 nVOC.
- Rieke (8;10) and Max (6;2) and their parents do also not realise different rhythmic patterns for each language. Likewise, the PVIs of the parents' respective L1 speech do not differ significantly.

An analysis of the speech data gathered from *[Family C]* produced the following results:

- Salome's (13;2) nVOC and rINT values in both languages are very similar to her parents' values. They are either located between her parents' L1 and L2 results or, in the case of German nVOC, slightly lower. Reuben's values tend to be slightly higher than his parents', but with the exception of German rINT are not significantly different.
- The siblings produce a very similar overall pattern which is furthermore almost identical to the pattern produced by their mother.
- In both L1 and L2 the two children are closer to the rhythmic pattern produced by their mother than to that produced by their father.
- Salome and her mother both produce significantly different nVOC patterns for German and English. Reuben's English and German nVOCs are near-significant. It can be supposed that the rhythms are still being teased apart, but this would have to be tested by an analysis of his speech at a later point.

The *comparison of child and adult* nVOC and rINT values shows that the younger children have a low vocalic PVI and a high intervocalic PVI and the older children have a comparatively higher nVOC and lower rINT. This seems to indicate that there is a general developmental trend which requires the children to increase vocalic variability (nVOC) and reduce intervocalic variability (rINT). That means that young children initially produce vowels of roughly equal length and then have to acquire the correct amount of vowel reduction. Similarly, consonant clusters are initially produced with a greater amount of durational variability, which is then gradually reduced.

A similar result with regard to the vocalic PVI in child speech has been found in a cross-linguistic study by Grabe, Gut, Post & Watson (1999). They concluded that children proceed from a less variable to an increasingly variable rhythmic pattern during the acquisition process. Unfortunately, they did not measure intervocalic interval data, so a comparison is not possible.

In their study they also concluded that French and German monolingual children have acquired the rhythmic structure of their language by age 4, whereas English children have not. In contrast, bilingual children do not seem to have successfully acquired either German or English rhythmic patterns until much later, at around age 11. Taking into account previous research on the acquisition of speech timing which suggests that, as the last durational feature to be acquired, consonant duration in unstressed position becomes adult-like only at age 10 (Kent 1976), it seems unlikely that the rhythm of a language can be adult-like before that, since rhythmic patterning of language is necessarily linked to segment duration. The performance of the bilinguals in this study seems to confirm the results of these earlier studies.

The *similarity of the siblings' speech* can be explained by the fact that they will be exposed to a very similar linguistic input. It can also be assumed that there is some kind of mutual or one-way influence of one sibling's speech to the other's.

The *disparity between bilingual and monolingual L1 patterns* demonstrates again that bilinguals cannot be regarded as two monolinguals in one (Grosjean 1998). At the same time this study provides evidence for a separation of rhythmic patterns. Bilingual children will differentiate between the phonetic and phonological structure of their languages if there are any differences in the linguistic input. Salome, the oldest child of [Family C] has indeed acquired two separate rhythmic patterns present in her mother's speech. Additionally, her younger brother Reuben appears to be in the process of acquiring these patterns. There is thus a strong connection between

linguistic input and acquisition, which has been previously remarked upon in the literature (e.g. Romaine 1995, 213-216). The other factor which seems to play a major part in acquisition is, quite unsurprisingly, the age of the children (e.g. Leather 1999, 10). Salome, who is almost two years older than her brother, realises separate patterns, whereas her brother has not developed two separate patterns yet.

The *closeness of bilingual patterns to L2 patterns* is very likely due to the fact that in both cases there will be some influence between the two linguistic systems. The difference between the two appears to be the degree to which the two patterns can be acquired and separated. It has to be borne in mind though that the L2 speakers in this study, in particular the L2 English of the mothers, is very advanced. Both the fathers and the mothers have acquired their second languages to a large degree by a process of immersion, which is closer to L1 acquisition than learning a language by means of instruction only (e.g. Heyd 1991, 13).

12 Conclusion

If children are exposed to different rhythmic patterns in the linguistic input they receive, they will realise different patterns by the time speech motor development has been completed. This is also the case if the rhythms are only distinguished with regard to fine-grained phonetic detail, as for German and English speech rhythm.

The acquisition of two different rhythmic patterns by bilinguals is dependent on age and linguistic input. The acquisition of rhythm in bilinguals appears to be only completed some time around age 11. This agrees roughly with previous research which indicates that development of speech timing concludes around age 10 (Kent 1976). The linguistic input is also crucial. The only children who display significantly different or near-significant rhythmic patterns for German and English are Salome and Reuben, whose mother's L1 and L2 productions differ significantly. Those of the other parents do not and thus their children will be unlikely to develop different patterns.

There is some evidence that the parents' speech patterns influence their children's speech to a certain degree. The children of [Family C] very much emulate the speech pattern of their mother. However, further research is required to confirm this finding.

Even if separate rhythmic patterns are realised, the rhythmic patterns of bilinguals are not necessarily identical to the monolingual patterns of the respective languages, although perceptually the bilinguals sound native-like even to trained ears. The results of this study indicate that bilingual speech is closer to L2 speech at least where immature, developing speech is concerned. This can be interpreted as evidence that the difference between L2 and bilingual speech is one of degree of attainment.

Finally, it is necessary to call for further research on rhythm, and specifically the PVI. This measure appears to be useful as an acoustic correlate of rhythm. However, it is far from clear what perceptual impression is created by any specific vocalic or intervocalic PVI value or any combinations thereof. It is also necessary to collect a greater database of normative values for languages and accents of languages.

References

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press

Adams, C. (1979). *English Speech Rhythm and the Foreign Learner*. The Hague: Mouton.

Allen G. & Hawkins, S. (1980). Phonological rhythm: definition and development. In: G H Yeni-Komshian, J F Kavanagh & CA Ferguson. *Child Phonology*. *Volume 1: Production*. Academic Press: New York. 227-256.

Bolinger, D. (1965). Pitch accent and sentence rhythm. In: Abe, I. & Kanekio, J. (eds.). *Forms of English: Accent, Morpheme, Order*. Cambridge/Mass.: Harvard University Press. 139-180.

Burnett, J. Lord. (1774). Of the Origin and Progress of Language.

Butcher, A. (1981). Aspects of the speech pause: phonetic correlates and communicative function, *Arbeitsberichte,* **15**. Institut für Phonetik,Universität Kiel.

Buxton, H. (1983). Temporal predictability in the perception of English speech. In: Cutler, A. & Ladd, D. R. *Prosody: Models & Measurements*. Berlin et al.: Springer-Verlag.

Classe, A. (1939). *The Rhythm of English Prose*. Oxford, Basil Blackwell.

Couper-Kuhlen, E. (1993). *English Speech Rhythm*. Amsterdam: John Benjamins.

Cruttenden, A. (1986). Intonation. Cambridge: Cambridge University Press.

Cutler, A. & Foss, D. J. (1977). On the role of sentence stress processing. *Language and Speech* **20**, 1-10.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalysed. *Journal of Phonetics* **11**, 51-69.

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the 11th International Congress of Phonetic Sciences*. Vol. 5, 447-450.

Delattre, P. (1965). *Comparing the Phonetic Features of English, French, German, and Spanish*. Heidelberg: Julius Groos Verlag.

Duez, D. (1982). Silent and non-silent pauses in three speech styles, *Language and Speech* **25**, 11-28.

Fantini, A. E. (1985). *Language Acquisition of a Bilingual Child: A Sociolinguistic Perspective*. Clevendon: Multilingual Matters.

Fox, A. (2000). *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals*. Oxford University Press: Oxford & New York.

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica* **38**,148-158.

Grabe, E. & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology* **7**.

Grabe, E. (1998). *Comparative Intonational Phonology: English and German*. MPI Press: Nijmegen.

Grabe, E., Gut, U., Post, B. & Watson, I. (1999). The acquisition of rhythm in English, French and German. *Proceedings of the Child Language Seminar 1999*.

Grabe, E., Post, B., & Watson, I. (1999). The acquisition of rhythm in English and French. *Proceedings of the 14th International Congress of Phonetic Sciences*. 1201-1204.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language & Cognition* **1**, 131-149.

Hakkarainen, H. (1995). *Phonetik des Deutschen*. München: W. Fink Verlag.

Halliday, M. A. K. (1967). *Intonation and Grammar in British English*. The Hague & Paris: Mouton.

Halliday, M.A. K. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.

Halliday, M.A.K. (1970). *A Course in Spoken English: Intonation*. Oxford: Oxford University Press.
Heyd, G. (1991). *Deutsch lehren*. Frankfurt: Diesterweg.
Kaltenbacher, E. (1997). German speech rhythm in L2 acquisition. In: Leather, J. & James, A. (eds.). *New Sounds '97. Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech.* Klagenfurt, University of Klagenfurt. 158-166.
Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies. *Journal of Speech and Hearing Research* **19**, 421-447.
Kohler, K. (1995) *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt Verlag.
Konopczynski, G. (1995). A developmental model of acquisition of rhythmic patterns: results from a cross-linguistic study. *Proceedings of the 13th ICPhS* **4**. 22-25.
Krakow, R. A. (1999). Physiological organization of syllables: a review. *Journal of Phonetics* **27**, 23-54.
Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.
Leather, J. (1999). Second-language speech research: an introduction. In: Leather, J. (ed.). *Phonological Issues in Language Learning*. Oxford: Blackwell. 1-58.
Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics* **5**, 253-263.
Low, E. L., Grabe, E. & Nolan, F. (2001). Quantitative characterisations of speech rhythm: 'syllable-timing' in Singapore English. *Language and Speech* **43**, 377-401.
Lyon, J. (1996). *Becoming Bilingual*. Clevendon: Multilingual Matters.
Nespor, M. (1990). On the rhythm parameter in phonology. In: Roca, I. (ed.) *Logical Issues in Language Acquisition*. Dordrecht: Foris. 157-175.
O'Connor, J. D. (1965). Perception of time intervals. *Progress Report* **2**, Phonetics Laboratory, University College London, 11-15.
O'Connor, J. D. (1968). The duration of the foot in relation to the number of component sound segments. *Progress Report* **3**, Phonetics Laboratory, University College London, 1-6.
O'Connor, J. D. (1973). *Phonetics*. Harmondsworth/UK: Penguin Books.
Peterson, G. E. & Lehiste, I. (1960). Duration of syllable nuclei. *Journal of the Acoustical Society of America* **32**, 693-703.
Pike, K. (1945). *Intonation of American English*. Ann Arbor Press: University of Michigan.
Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265-292.
Romaine, S. (1995). *Bilingualism*. 2nd edition. Oxford & Malden/MA: Blackwell.
Scott, D. R., Isard, S. D. & De Boysson-Bardies, B. (1985). Perceptual isochrony in English and in French. *Journal of Phonetics* **13**, 155-162.
Shen, Y. & Peterson, G. G. (1962). Isochronism in English., *Occasional Papers* **9**, University of Buffalo, Studies in Linguistics, 1-36.
Steele, J. (1779). *Prosodia Rationalis, or An Essay towards Establishing the Melody and Measure of Speech, to be Expressed and Perpetuated by Peculiar Symbols*. 2nd edition amended and enlarged. London: J. Nichols.
Stetson, R.H. (1951) *Motor Phonetics,* 2nd ed., Amsterdam: North-Holland.

Tingley, B. M. & Allen, G. D. (1975). Development of speech timing control in children. *Child Development* **46**, 186-194.

Uldall, E. T. (1971). Isochronous stresses in R. P. In: Hammerich, L. L., Jakobson, R. & Zwirner, E. (eds.). *Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen*. Copenhagen: Akademisk Forlag.

Nicole Whitworth
Clinical Language Sciences
Leeds Metropolitan University
City Campus
Leeds, UK
LS 1 3HE

N.Whitworth@lmu.ac.uk