An empirical study of the perception of language rhythm

Franck Ramus, Emmanuel Dupoux, Renate Zangl, and Jacques Mehler
Laboratoire de Sciences Cognitives et Psycholinguistique (EHESS/CNRS)

Draft of June 7, 2000

Linguists have traditionally classified languages into three rhythm classes, namely stress-timed, syllable-timed and mora-timed languages. However, this classification has remained controversial for various reasons: the search for reliable acoustic cues to the different rhythm types has long remained elusive; some languages are claimed to belong to none of the three classes; and no perceptual study has bolstered the notion. However, Ramus, Nespor & Mehler (1999), *Cognition 73*, 265-292, have recently proposed an acoustic/phonetic model of the different types of linguistic rhythm, and of their categorization as such by listeners. Their simulations make predictions as to which languages can be discriminated on the basis of their rhythm. Here, we present perceptual experiments that directly test the notion of rhythm classes, the simulations' predictions and the question of intermediate languages. Language discrimination experiments were run using a speech resynthesis technique to ensure that only rhythmical cues are available to the subjects. Languages investigated are English, Spanish, Catalan and Polish. Discrimination results are compatible with the rhythm class hypothesis, but Polish rhythm seems to be different from any other language studied and thus may constitute a new rhythm class. A revised version of the rhythm perception model is proposed to accommodate these findings and more simulations are run to generate new predictions.
Key-words: speech rhythm, prosody, language discrimination, speech perception, stress-timing, syllable-timing.

The perception of language rhythm has been a subject of interest among linguists for decades. This interest comes from the observation that different languages give rise to the perception of different types of rhythm. For instance, Lloyd James (1940) noted that English has a rhythm similar to that of a Morse code, while Spanish rhythm is closer to that of a machine-gun. Pike (1945) attributed this difference to the fact that in English rhythm is due to the recurrence of stresses, while in Spanish it is due to the recurrence of syllables, hence the terminology "stress-timed" and "syllable-timed" languages. Abercrombie (1967) further claimed that all languages should fall into one of these categories, and that rhythmical structure was based on the isochrony of the corresponding rhythmical units, that is, the isochrony of interstress intervals for the former category and the isochrony of syllables for the latter. However, measurements in the speech signal have failed to provide empirical support for this "isochrony theory" (Bolinger, 1965; Roach, 1982; Dauer, 1983).

Dasher and Bolinger (1982) and Dauer (1983) have advocated another view of speech rhythm, according to which the different types of rhythm are due to different sets of phonological properties. Indeed, stress-timed languages authorize complex syllables and have vowel reduction, while syllable-timed languages only have simpler syllables and no vowel reduction. This view raises the possibility that different combinations of these phonological properties might give rise to rhythms that are neither stress- nor syllable-timed. Nespor (1990) has indeed argued that some languages present such combinations: Catalan has simple syllables, like syllable-timed languages, and vowel reduction, like stress-timed languages; Polish, on the other hand, has complex syllables and no vowel reduction at normal speech rates. Nespor (1990) thus concludes that these languages should have intermediate types of rhythm. At that stage, what is missing is primary empirical evidence. Do human listeners categorically perceive different types of rhythm? If so, how would they classify languages such as Catalan and Polish? And what auditory cues support rhythm perception? These are the questions the present study aims to address.

Insights into the perception of speech rhythm by adult listeners can also have implications for the study of language acquisition. Experts in this field are indeed unanimous in thinking that rhythm plays an essential role in the first stages of language acquisition (Cutler, 1994; Mehler, Dupoux, Nazzi, & Dehaene-Lambertz, 1996; Morgan, 1996;

This work was supported by grants from the Dlgation Gnrale pour l'Armement to F.R. and J.M., and from FWF (project # J-1646/SPR) to R.Z.. We thank Thierry Fourneyron and Sid Kouider for their help testing the subjects, Núria Sebastián-Galls and Laura Bosch for recording the Catalan and Spanish material, Christophe Pallier for discussions and advice, and Anne Christophe, Marina Nespor and Derek Mitchell for comments on a previous version.

Renate Zangl is now at Center for Research in Language, University of California, San Diego.

Correspondence should be addressed to Franck Ramus, now at Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR, Great-Britain. E-mail: f.ramus@ucl.ac.uk.

Jusczyk, 1998), possibly bootstrapping the task of word segmentation. Alternatively, speech rhythm has been proposed as a cue to syllable structure (Ramus, Nespor, & Mehler, 1999). A better understanding of speech rhythm perception in adults, supplemented by studies conducted on infants, may provide an empirical basis to these hypotheses.

Auditory cues to rhythm perception

Before we tackle empirical studies of speech rhythm perception, it may be relevant to ask whether measurable regularities or properties of the speech signal can predict listeners' classification of rhythm types. As we have mentioned above, measurements inspired by the isochrony theory have had little success. However, building upon the phonological account of rhythm (Dasher & Bolinger, 1982; Dauer, 1983), Ramus et al. (1999) have shown that the type of rhythm of a given language can be deduced from the duration of its vocalic and consonantal intervals. This finding not only sheds light on a point of linguistic typology, but also suggests a model of how speech rhythm may actually be perceived. From this model it is possible to make predictions as to which languages should be perceived as having the same rhythm and which should be perceived as having different types of rhythm. Since these predictions bear directly on our present experiments, we now present the model of Ramus et al. (1999) and the corresponding simulations.

Measurements

Measurements of the duration of consonantal and vocalic intervals were made in eight languages (English, Dutch, Polish, French, Spanish, Italian, Catalan and Japanese). Consonantal intervals span sequences of consecutive consonants, while vocalic intervals span sequences of consecutive vowels. These measurements thus assume a coarse consonant/vowel segmentation of speech, but not a detailed phonetic segmentation[1]. Three variables are computed from the measurements, taking one value per sentence: $\%V$, the percentage of the sentence's duration taken up by vowels; ΔV, the standard deviation of vocalic intervals within the sentence; and ΔC, the standard deviation of consonantal intervals. Ramus et al. (1999) found that the eight languages studied cluster in three groups along the $\%V$ and ΔC dimensions, and that these groups correspond to three distinct rhythm types[2].

The model

The above measurements suggest a straightforward model of how the human auditory system extracts rhythm type from speech. The model relies on the following prerequisite abilities: (a) discrimination of consonants vs. vowels, (b) evaluation of time intervals, (c) computation of simple statistics. It takes the following steps:

1. Segment the speech sample into consonantal and vocalic intervals;
2. Measure their respective durations;
3. Compute $\%V$, ΔV, and ΔC.

Rhythm type follows from the values obtained.

Simulations

From this model it is possible to make predictions as to which languages are discriminable on the basis of their rhythm. However, specific predictions depend on the particular task that is used to assess discrimination. The following simulations are modeled on the kind of language discrimination experiments previously run by Ramus and Mehler (1999). In these experiments, subjects are trained to discriminate between 10 sentences of language L1 and 10 sentences of language L2, uttered by 2 speakers per language. In a subsequent test phase, subjects are tested on their ability to correctly classify 10 new sentences of each language, uttered by 2 new speakers per language.

For any given pair of languages, a simulation was run by performing a logistic regression taking language as categorical variable, and $\%V$ as predictor variable, on 20 training sentences. The result threshold value for $\%V$ was then used to classify the test sentences. Predictions are reproduced in Table 1.

Table 1
Language discrimination simulations based on $\%V$. Scores are percentages of test sentences correctly classified. Chance level is 50%. Cases in italics reflect between-class language pairs.

	Eng.	Dut.	Pol.	Fre.	Ita.	Cat.	Spa.
Dut.	57.5						
Pol.	50	57.5					
Fre.	*60*	*60*	*65*				
Ita.	*65*	*62.5*	*65*	55			
Cat.	*65*	*62.5*	*65*	57.5	35		
Spa.	*62.5*	*57.5*	*62.5*	50	50	37.5	
Jap.	*92.5*	*92.5*	*95*	*90*	*90*	*87.5*	*95*

The simulations thus predict fair classification of sentences ($\geq 60\%$) between languages with different types of rhythm and chance performance ($< 60\%$) for languages with the same type of rhythm[3]. These predictions are thus entirely consistent with what Mehler et al. (1996) and Nazzi,

[1] This was justified by evidence that newborns are particularly sensitive to vocalic intervals of speech (Mehler et al., 1996). Recent data suggest that consonants and vowels may indeed be processed differently by the brain (Caramazza, Chialant, Capasso, & Miceli, 2000).

[2] English, Dutch and Polish form the stress-timed group, French, Spanish, Italian and Catalan the syllable-timed group, and Japanese is the only member of the mora-timed group. This latter group, although not recognized in the earliest descriptions of language rhythm, brings together languages whose rhythm is due to the mora, a sub-syllabic unit (Ladefoged, 1975). Within the phonological description of rhythm, mora-timed languages are those with the simplest syllabic structure.

[3] With the exception of the Spanish/Dutch pair, for which a 57.5% score is predicted, although they are respectively syllable-

Bertoncini, and Mehler (1998) called the rhythm class hypothesis, insofar as Polish is considered as stress-timed, and Catalan as syllable-timed.

Thus, the model set forth by Ramus et al. (1999) makes particular predictions concerning Nespor's (1990) intermediate languages. A study of the perception of speech rhythm including Catalan and Polish may therefore provide an empirical test of both Nespor's claim and Ramus et al.'s model. This is the purpose of the following sections.

Material

Languages

Consistent with the predictions in Table 1, Ramus and Mehler (1999) showed that English and Japanese rhythms can be discriminated. Among the 27 remaining language pairs in Table 1, the selection we chose to test aims at evaluating Nespor's (1990) hypothesis about intermediate languages. We chose to compare Catalan and Polish with two reference languages, English for the stress-timed type and Spanish for the syllable-timed[4].

Sentences

Sentences were selected from a multi-language corpus initially constituted by Nazzi (1997; Nazzi et al., 1998) and extended for the present study (Polish and Catalan). In this corpus, 54 sentences of each language were read by four female native-speakers. We selected 5 sentences per speaker, thus constituting a set of 20 sentences per language. Sentences were selected in such a way as to minimize differences across languages: they were thoroughly matched in number of syllables (from 15 to 19 syllables per sentence, with an average of 17), and had comparable durations and average fundamental frequencies[5] (see Table 2).

Since an ANOVA showed that there were significant differences in duration between Polish and the other languages, and in fundamental frequency between English, Spanish and Polish, these differences were compensated for through the resynthesis process described in the following section.

Resynthesis

In order to ensure that only the relevant cues are made available to the subjects in language discrimination experiments, Ramus and Mehler (1999) developed a speech resynthesis technique, allowing to selectively degrade or preserve the phonological and prosodic properties of speech. Basically, each sentence is resynthesized by feeding its phonetic transcription, the duration of each phoneme and the fundamental frequency curve into an adequate speech synthesizer. Here, we use MBROLA (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996)[6], which performs the synthesis through concatenation of diphones, using a French diphone database. Phonological properties of the synthesized sentences are manipulated by modifying the input data to the synthesis. Here we only use two of the four types of manipulation documented in Ramus and Mehler (1999):

sasasa: all consonants are replaced by /s/ and all vowels are replaced by /a/. The fundamental frequency is preserved. Therefore, all lexical, phonetic and phonotactic information is eliminated, preserving only the prosodic information, that is rhythm and intonation.

flat sasasa: the phonemes are replaced in the same manner as for the *sasasa*, but the original fundamental frequency of the sentences is ignored, and replaced by a constant one at 230 Hz. This way, the synthesized stimuli can only convey the rhythmical properties of the original sentences.

Sample stimuli can be heard on
`http://www.ehess.fr/centres/lscp/persons/ramus/resynth/ecou`

Incidentally, the resynthesis process gives total control over the duration and the average fundamental frequency of the sentences. We used this feature in order to reduce the differences found across languages, and eliminate the possibility that subjects might discriminate languages on the basis of such artifactual cues as sentence duration or pitch. We therefore multiplied all fundamental frequency values of the Spanish sentences by a factor of 1.12. As regards Polish, a closer scrutiny revealed that unlike the other languages, there were important differences between the four speakers. To reduce these differences as well, different factors were applied to each speaker: sentences' durations were multiplied respectively by 1, 0.92, 0.9 and 0.93, and fundamental frequency values by 1.07, 1.09, 1.17 and 0.91. Additional measurements performed on the resynthesized sentences ensured that no significant differences in F_0 or duration remained across the four languages. Informal listening revealed no perceptible difference or artifact as a result of certain sentences having their duration or pitch modified (the factors used were indeed quite close to 1).

Language discrimination on the basis of rhythm and intonation

Although only the *flat sasasa* version of the sentences is relevant to the study of speech rhythm perception, we conducted a number of preliminary studies using the *sasasa* version, hoping to increase our chances of observing discrimination effects.

Exp. 1-2: Calibration and optimization of the discrimination task

Ramus and Mehler (1999) used a categorization task (X), where subjects were trained to correctly classify the two languages on half the sentences, and then tested on the other

and stress-timed.

[4] Catalan and Polish are obviously not mora-timed languages.

[5] Fundamental frequency was extracted at intervals of 5 ms using the Bliss software. Average fundamental frequency was computed as the average of all non-zero fundamental frequency values for each sentence.

[6] MBROLA is freely available from `http://tcts.fpms.ac.be/synthesis/mbrola.html`.

Table 2
Average duration and fundamental frequency across sentences of the different languages. Standard deviation in parentheses.

	English	Spanish	Polish	Catalan
Duration (ms)	2852 (207)	2840 (243)	3034 (330)	2856 (285)
Average F_0 (Hz)	230 (16)	206 (17)	219 (22)	[a]

[a] This was not measured, since Catalan sentences were not used in any experiment involving fundamental frequency (see below).

half. Although the languages studied (English and Japanese) were maximally different, at least as far as rhythm is concerned, classification scores in the training phase were never dramatically high (at best 0.72 for the A' score). Thus one may fear that the discrimination of finer differences may fail to be observed, and feel the need to use a more sensitive task.

The X task involves finding the relevant cues and forming categories for the two languages over the training sentences, and generalizing this categorization to the test sentences. We now propose to evaluate a potentially less demanding task, the odd-ball task (AAX). Here, two sentences of the same language are played as a context, then a third sentence is played either in the same language or in a different one, and the subject should respond Same or Different. Thus this task requires only the immediate comparison of series of three sentences, with no training and no generalization.

Method. We used the 20 English and the 20 Spanish sentences resynthesized in the *sasasa* manner as described above. All experimental protocols were programmed on a PC using the EXPE language (Pallier, Dupoux, & Jeannin, 1997)[7].

Experiment 1 uses the X task following the same protocol as in Ramus and Mehler (1999). First the instructions are displayed on the screen, informing subjects that they are to identify two exotic languages, Sahatu and Moltec. Sentences are divided into a training set (2 speakers per language) and a test set (2 other speakers per language); they are played through a ProAudio Spectrum sound card, and heard through headphones. In the training phase, subjects first listen to two example sentences, then to the 20 training sentences one at a time, and have to answer S for Sahatu, or M for Moltec. They receive immediate feedback on their answer, and then have the opportunity to listen to the sentence again. A score of 70% correct responses or more allows them to switch to the test phase. Otherwise, they undergo the training phase again up to a maximum of three times, after which they switch to the test phase regardless of their score. In the test phase, subjects listen to the 20 test sentences one at a time, answer S or M and get immediate feedback as in the training phase. Each test sentence is heard only once.

Experiment 2 uses the AAX task as follows. The two sets of sentences defined for Exp. 1 now constitute a context set (AA in AAX) and a test set (X in AAX). The session includes two blocks of 20 trials each, during which the context language is held constant. At the beginning of each block, subjects are told which language (Sahatu or Moltec) is the context language. Blocks were counterbalanced across subjects. For any given trial, the two context sentences are drawn randomly from the context set, subject to the constraint that they are uttered by two different speakers. The test sentence is also drawn randomly from the test set, and each test sentence is played only once per block. Within a trial, sentences were played with an inter-stimulus-interval of 500 ms. Subject were then required to press a key to indicate whether they thought the third sentence was expressed in the same language as the first two.

Participants. Thirty-two students participated, 16 in each experiment. There were 22 men and 10 women, with a mean age of 21. They were all French native speakers. They voluntarily participated without any payment, and were tested in their own room, using a portable PC.

Results. In Exp. 1, only the results of the test phase are taken into account. Hit and false alarm rates are computed for each subject as follows: in Exp. 1, hit rate is the proportion of Spanish sentences correctly identified, and false alarm rate is the proportion of English sentences incorrectly labeled as Spanish; in Exp. 2, hit rate is the proportion of correct "same" trials and false alarm rate is the proportion of incorrect "different" trials. In order to take into account possible response biases, and following Ramus and Mehler (1999), hit and false alarm rates are converted into A' discrimination scores[8]. Table 3 presents average A' scores for the two experiments.

Table 3
English-Spanish discrimination using sasasa *stimuli.*

	A'	St. Dev.	p
Exp. 1 (X)	0.62	0.19	0.024
Exp. 2 (AAX)	0.61	0.09	< 0.001

The distributions of A' scores can be considered normal for both experiments ($p > 0.20$ with a Lilliefors test[9]). Av-

[7] EXPE is freely available from `http://www.ehess.fr/centres/lscp/expe/expe.html`.

[8] A' is a variant of d' which requires less stringent assumptions. Let H be the hit rate and F the false alarm rate:

$$\text{If } H \geq F, \text{ then } A' = \frac{1}{2} + \frac{(H-F)(1+H-F)}{4H(1-F)}$$
$$\text{if } H < F, \text{ then } A' = \frac{1}{2} - \frac{(F-H)(1+F-H)}{4F(1-H)}$$

See Snodgrass and Corwin (1988) for further details.

[9] This is a variant of the Kolmogorov-Smirnov normality test which does not assume the mean and the variance to be known.

erage A' scores are significantly above 0.5 (chance level), both for Exp. 1 ($t(15) = 2.52$, $p = 0.024$), and for Exp. 2 ($t(15) = 4.45$, $p < 0.001$). Thus, in both experiments, subjects managed to discriminate between the two languages, and achieved similar scores. However, subjects' scores are more variable in Exp. 1 than in Exp. 2, as confirmed by a Levene's test for equality of variances (p=0.12, non significant trend). This explains the difference in significance between the two t-tests, and suggests that the AAX procedure might have higher statistical power.

Discussion. These experiments show that English and Spanish can be discriminated on the basis of their prosodic properties. *Flat sasasa* stimuli are now necessary to assess whether rhythm is sufficient for the discrimination.

The lesser variance of A' scores found using the AAX task incites us to prefer this task for future experiments, where discrimination effects may be more difficult to detect. Although Levene's test was not significant, it remains possible that it is due to chance or to another factor. However, since we have no other reason to prefer one task over the other, we will now adopt the AAX task for all the other experiments.

Exp. 3-4: English-Polish and Spanish-Polish

Method. Here we use the English, Spanish and Polish sentences described above, resynthesized in the *sasasa* manner. Experiment 3 tests discrimination between English and Polish, and Experiment 4 discrimination between Spanish and Polish. The procedure is the same as for the Exp. 2, using the AAX task.

Participants. Thirty-two students participated, 16 in each experiment. There were 23 men and 9 women, with a mean age of 22. They were recruited and tested as in Exp. 1-2.

Results. Table 4 gives the mean A' scores for Exp. 3-4, together with Exp. 2 for comparison. The distribution of A' scores can be considered normal for both experiments ($p > 0.2$ with a Lilliefors test). A' scores are significantly greater than 0.5 in Exp. 4 ($t(15) = 4.14$, $p = 0.001$), but not in Exp. 3 ($t(15) = 1.183$, $p = 0.25$). Thus, subjects seem to be able to discriminate between Spanish and Polish, but not English and Polish.

Table 4
English-Spanish, English-Polish and Spanish-Polish discrimination using sasasa *stimuli.*

	A'	St. Dev.	p
Exp. 2: English-Spanish	0.61	0.09	< 0.001
Exp. 3: English-Polish	0.55	0.18	0.25
Exp. 4: Spanish-Polish	0.65	0.15	0.001

For a better visualization of the data, Figure 1 shows histograms of A' scores for the three experiments. The solid vertical line separates scores ≤ 0.5 from scores > 0.5[10]. Dashed vertical lines delimit the interval of scores which are not significantly different from chance[11]. It should be noted that low scores do not reflect discrimination and mislabeling: this is a same/different task, with marked keys indicating the meaning of the response (P for "Pareil" and D for "Diffrent"), and with feedback being provided after each trial. Low scores thus merely reflect "bad luck" or perseveration in a wrong strategy.

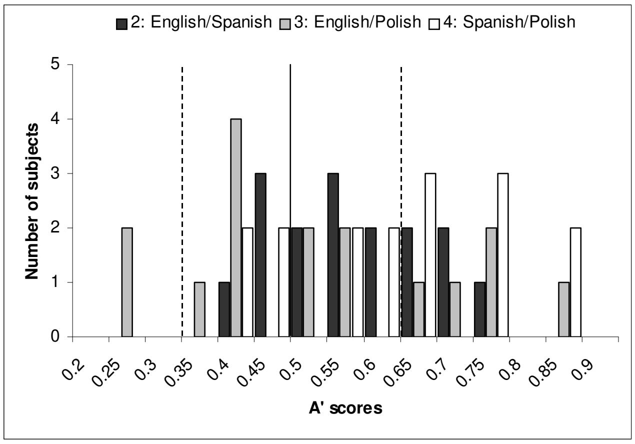

Figure 1. Histogram of A' scores for Exp. 2, 3, and 4.

Our results suggest that Polish should perhaps be considered a stress-timed language like English. It is however important to assess more directly this hypothesis using the *flat sasasa* stimuli, which will preserve rhythm only. This is the purpose of the next section.

Language discrimination on the basis of rhythm only

Exp. 5-7: English-Spanish, English-Polish and Spanish-Polish

Method. We used the same English, Spanish and Polish sentences as in experiments 1-4, but this time they were resynthesized in the *flat sasasa* manner. Their duration was corrected as described above, and the fundamental frequency needed no correction, since it was made constant. Experiments were run following exactly the same procedure as for Exp. 2-4.

Participants. Forty-eight students participated, 16 in each experiment. There were 34 men and 14 women, with a mean age of 22. They were recruited and tested under the same conditions as for Exp. 2-4.

Results. Table 5 gives the mean A' scores for Exp. 5-7, and Figure 2 their distribution.

According to the Lilliefors test, the distribution of A' scores cannot be considered normal for Exp. 5 ($p = 0.01$), and for Exp. 6 the non-normality is nearly significant ($p =$

[10] Intervals are of the $]n, n + 0.05]$ kind.

[11] Since there are 40 observations per subject, individual scores need to be greater than 0.66 or lower than 0.33 to be significantly different from 0.5, according to a 2-tail binomial test.

Table 5
English-Spanish, English-Polish and Spanish-Polish discrimination using flat sasasa *stimuli.*

	A'	St. Dev.	p
Exp. 5: English-Spanish	0.65	0.13	0.01^a
Exp. 6: English-Polish	0.61	0.14	0.006
Exp. 7: Spanish-Polish	0.74	0.08	< 0.001

[a] This level of significance was obtained through a binomial test, and is thus not directly comparable to the others.

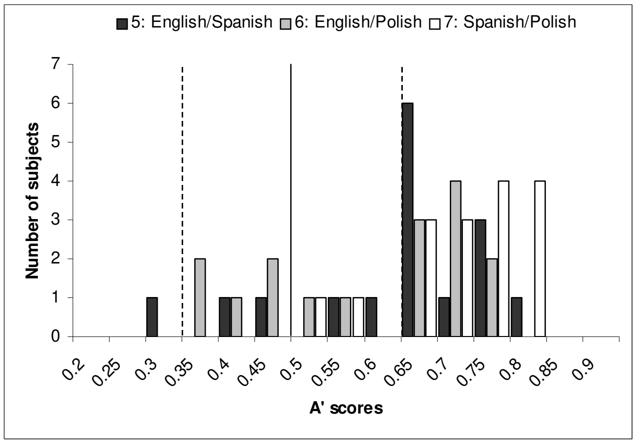

Figure 2. Histogram of A' scores for Exp. 5, 6, and 7.

0.056); this is not the case, however, for Exp. 7 ($p = 0.15$). Since the assumption of normality required for a t-test is not met in the case of Exp. 5, we use instead a binomial test comparing the number of subjects with a score > 50 and those with a score ≤ 50. The standard t-test is used for Exp. 6 and 7. Obviously, the significance levels will not be comparable across experiments.

For all three experiments, the A' scores are significantly above 0.5: $p = 0.01$ with a binomial test for Exp. 5, $t(15) = 3.24$, $p = 0.006$ (and $p = 0.04$ with a binomial test) for Exp. 6, and $t(15) = 11.59$, $p < 0.001$ for Exp. 7. Thus, subjects seem to be able to discriminate English from Spanish, English from Polish, and Spanish from Polish. Since *flat sasasa* stimuli were used, we may conclude that they did so only on the basis of the languages' rhythmical properties. The English-Polish discrimination may seem in contradiction with the result found in Exp. 3, where these two languages were not discriminated in their *sasasa* version. This actually suggests that the presence of multiple cues is not necessarily beneficial, but rather may distract subjects from the most relevant one, i.e. rhythm (this had already been noted by Ramus & Mehler, 1999). The English-Polish discrimination is nevertheless unambiguous when only rhythm is available.

Discussion. The results obtained are consistent with Nespor's (1990) hypothesis that Polish has rhythmical properties unlike those of both syllable-timed and stress-timed languages, and thus that it might belong to an intermediate class of languages. English-Polish discrimination was not predicted by Ramus et al.'s (1999) simulations based on the variable $\%V$; the model may thus gain in accuracy by incorporating the variable ΔV. In terms of phonological properties, this suggests that syllabic complexity does not fully account for speech rhythm. Indeed, English and Polish have similar syllabic complexities, and their discrimination must therefore be based on another property, for instance vowel reduction, which is present in English but not in Polish. Likewise, Spanish and Catalan have similar syllabic structure, but only Catalan has vowel reduction. Investigating the discrimination between Catalan and the other languages should now shed more light on this issue.

Exp. 8-10: Spanish-Catalan, English-Catalan and Polish-Catalan

The next three experiments are designed to test the rhythmical status of Catalan with respect to stress-timed languages (English), syllable-timed languages (Spanish), and a supposedly intermediate language (Polish).

Method. The English, Polish and Catalan sentences are those described in the *Material* section above. However, the Spanish sentences previously used have been criticized by Spanish colleagues, arguing that they were not representative of the mainstream Castillan dialect. Although there is no reason to assume that this may have influenced the results of the preceding experiments, we decided to have the Spanish corpus re-recorded by four unequivocal native Castillan speakers. 20 sentences were selected as before. Their mean duration was 2928 ms ± 214, not significantly different from that of the other languages ($F(3,76) < 1$)[12]. All sentences were resynthesized in the *flat sasasa* manner, bringing their F_0 at a constant 230 Hz. The procedure is the same as for Exp. 2-7.

Participants. Forty-eight students participated, 16 in each experiment. There were 21 men and 27 women, with a mean age of 24. They were recruited and tested under the same conditions as for Exp. 2-7.

Results. Table 6 gives the mean A' scores for Exp. 8-10, and Figure 3 their distribution.

The distribution of A' scores in Exp. 8, 9, 10 can be considered as normal (respectively $p > 0.2$, $p > 0.2$ and $p = 0.066$). In experiment 8, subjects failed to discriminate between Spanish and Catalan ($t(15) = -1.1$, $p = 0.3$). In experiments 9 and 10, A' scores are above chance level,

[12] In addition, a preliminary experiment revealed that discrimination between Spanish and Catalan was possible, but likely due to an artifact: 11 Spanish sentences out of 20 contained a silence, whereas only 2 Catalan sentences out of 20 did. To reduce the risk that subjects may employ a strategy based on the detection of silences, we removed all silences during resynthesis for the Spanish and Catalan sentences. We checked a posteriori that our English, Polish and former Spanish sentences did not present this problem.

Table 6
Spanish-Catalan, English-Catalan and Polish-Catalan discrimination using flat sasasa *stimuli.*

	A'	St. Dev.	p
Exp. 8: Spanish-Catalan	0.46	0.13	0.3
Exp. 9: English-Catalan	0.57	0.14	0.06
Exp. 10: Polish-Catalan	0.57	0.16	0.09

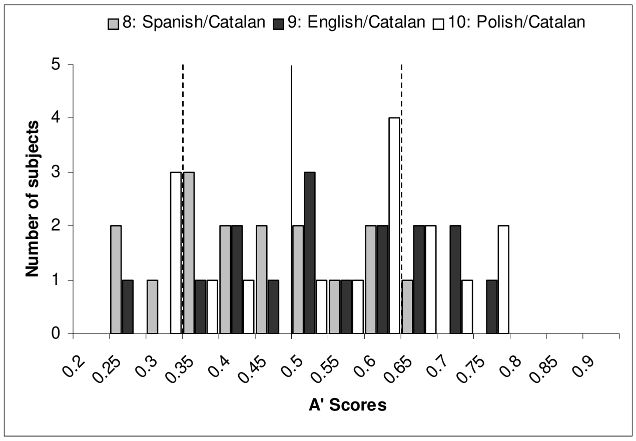

Figure 3. Histogram of A' scores for Exp. 8, 9, and 10.

with marginal significance: $t(15) = 2$, $p = 0.06$ for English-Catalan, $t(15) = 1.81$, $p = 0.09$ for Polish-Catalan.

It can be argued that although the t-tests we are using are 2-tail, the hypothesis being tested is that subjects have scores strictly greater than 0.5. In such circumstances, it is more appropriate to use a 1-tail test, even though this is not customary. It can thus be considered that the actual level of significance is $\alpha = 0.10$, and that discrimination was achieved in Exp. 9 and 10. It can also be noticed that the low mean A' scores are partly due to a few subjects scoring surprisingly well below chance level, although not as low as might be predicted by the null hypothesis considering the highest scores. In both cases, the distribution is clearly shifted towards the high scores. Finally, let us note that A' scores are significantly higher in Exp. 9 and 10 than in Exp. 8: respectively $t(30) = 2.2$, $p = 0.035$ and $t(30) = 2.1$, $p = 0.045$. Comparable significance is found using Mann-Whitney rank tests. The above arguments thus lead us to conclude that the low significance of tests for Exp. 9 and 10 is mainly a matter of statistical power, and that both the English-Catalan and the Polish-Catalan pairs are discriminable on the basis of rhythm, whereas the Spanish-Catalan pair is not.

Discussion. Unlike what we found with Polish, our results suggest that Catalan is not an intermediate language, but rather that it is a standard syllable-timed language, like Spanish. This is indicated by the fact that its rhythm is not discriminable from that of Spanish, although it is discriminable from both that of English, a stress-timed language, and Polish, an intermediate language. This result is consistent with Ramus et al.'s (1999) which predicted indeed that Catalan should cluster with syllable-timed language (see Table 1). However, the model was challenged by the results obtained on Polish, and it thus remains to be seen whether it can be adjusted to fit with the whole set of empirical data we have gathered.

Bimodality of the scores' distribution

Examination of Figures 1, 2 and 3 may suggest that the distributions of scores may be bimodal, composed of a normal distribution centered on 0.5, and another one centered on a higher score ($\simeq 0.75$). This would be compatible with the observation that certain subjects easily converge towards the right cues and strategy, whereas other subjects never do. Although the normality hypothesis was rejected only in the case of Exp. 5, it may be reassessed by looking at the distribution of scores over all the experiments.

Figure 4 presents the distribution of A' scores for all experiments (2-10) employing the AAX task. Visual examination, as well as a Lilliefors test show that this distribution is strongly non normal ($p = 0.001$), suggesting that the subjects tested do not have homogeneous performances. Figure 5 illustrates our bimodality hypothesis, by showing how the overall distribution can be seen as the sum of two normal distributions.

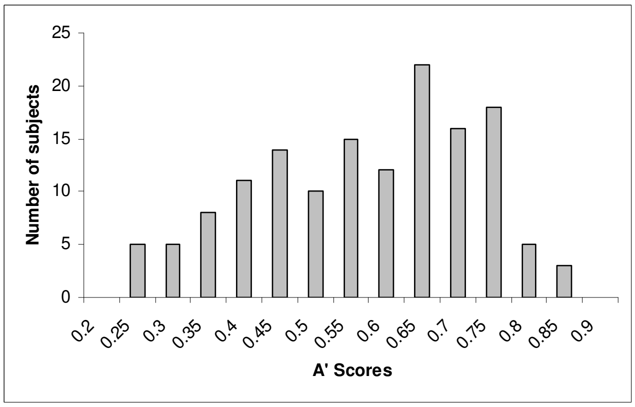

Figure 4. Histogram of A' scores for Exp. 2-10.

The fact that a relatively high proportion of subjects never find the right cue or strategy, together with the somewhat low level of average A' scores, confirms that the language discrimination task is a difficult one. It demands sustained concentration and attention. Whenever languages are harder to discriminate, the proportion of subjects failing the task increases, as illustrated by Exp. 9 and 10, and the average A' score decreases accordingly. Nevertheless, the difficulty of the task for certain pairs of languages should not mask the fact that it is *feasible*, as indicated by the high performance ($> 70\%$) of a number of subjects. Of all the experiments we have run, only Exp. 8 (Spanish-Catalan) is characterized by the complete absence of an upper mode for its distribution of scores, and can thus be declared unfeasible.

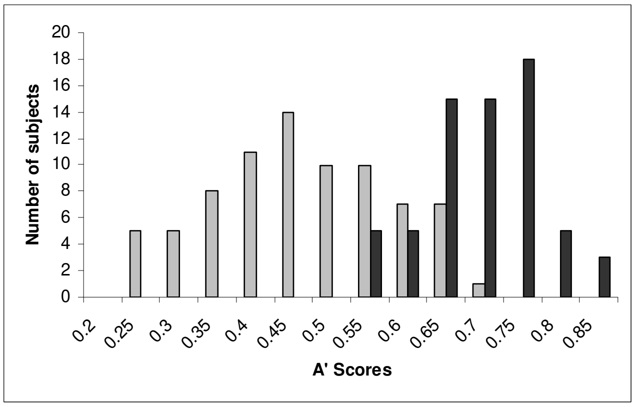

Figure 5. Breakdown of A' scores into two normal distributions (Exp. 2-10).

Adjusting the model to the new empirical data

We now possess empirical data concerning rhythm-based language discrimination. The results obtained are compatible with the predictions shown in Table 1, with the exception of the English-Polish pair. This is not a minor exception, however: it questions the standard rhythm classes as described in the linguistic literature. Since simulations by Ramus et al. (1999) were based on the variable $\%V$ with the explicit aim of fitting these rhythm classes, it is not surprising that the English-Polish discrimination was not predicted.

The question now is whether the model can be adjusted to fit the current pattern of data. Two variables, ΔV and ΔC, remain available to conduct new simulations. Ramus et al. (1999) had already noticed that a logistic regression based on ΔV predicted the English-Polish discrimination. It thus seems sensible to include this variable in the model. There are several ways to do this. One option is to base the logistic regressions exclusively on ΔV. However, it is predictable that this will not yield the expected pattern. Indeed, apart from the English-Polish pair, our data are entirely consistent with the standard rhythm classes, but ΔV alone does not account for these classes (see for example Table 1 or Figure 2 in Ramus et al., 1999). As a consequence, it seems preferable to use both variables.

Simulation 1: Language discrimination based on $\%V$ and ΔV

Here and in the following section, we use the same measurements of the duration of consonantal and vocalic intervals in eight languages as Ramus et al. (1999). The procedure used is again a logistic regression (cf. the *Simulations* section), which can be run as a multivariate procedure. Thus, we ran two[13] logistic regressions for each of the 28 pairs of languages considered. Input data now consist in ($\%V$, ΔV, L) triples, where L stands for one of the two languages considered. The regression reveals the threshold value of a linear combination of $\%V$ and ΔV that best discriminates between the two languages. Predicted percentages of correct responses are shown in Table 7. To reflect the experimental results obtained, Polish is now shown as an independent rhythm class.

The pattern of results obtained is not completely satisfactory: first, although the English-Polish discrimination is correctly predicted, the Polish-Catalan and Polish-Spanish ones are not, which is in direct contradiction with our data; second, the predicted discriminations between English and Dutch on the one hand, and Spanish and Italian on the other unnecessarily challenge the standard rhythm classes. It thus seems that the variable ΔV has had too great an influence on this simulation. Introducing ΔC, which is quite strongly correlated with $\%V$, may help produce more accurate predictions.

Simulation 2: Language discrimination based on $\%V$, ΔV and ΔC

Here again, we use logistic regressions, this time based on ($\%V$, ΔV, ΔC, L) quadruples. Besides the additional variable, the simulations are run exactly as before. Predicted percentages of correct responses are shown in Table 8.

The present predictions are compatible with the behavioral data obtained: discrimination of at least 60% of the sentences is predicted between English and Catalan, Spanish, Japanese and Polish, between Polish and Spanish and Polish and Catalan, but no discrimination at all is predicted between Spanish and Catalan. Moreover, scores below 60% are predicted within class (English-Dutch, and within the syllable-timed class, with the exception of Italian/Spanish at 60%), and scores equal or above 60% are predicted between classes (between syllable- and stress-timed languages, and between Japanese and the others). Thus, our simulations are consistent with both the empirical evidence and the standard rhythm classes. Moreover, Polish is predicted to be discriminable from all the other languages, and thus seems to form a rhythm class of its own.

Discussion

Although our empirical data didn't seem to fit with the predictions of Ramus et al.'s (1999) speech rhythm perception model based on $\%V$, new simulations show that the introduction of the other two variables is sufficient to accommodate the new findings.

We are now led to reconsider the model as follows: we assume that listeners segment speech into vocalic and consonantal intervals, compute statistics over their respective durations in the form of the variables $\%V$, ΔV and ΔC, and represent rhythm in the tri-dimensional space thereby defined. Language discrimination experiments using *flat sasasa* stimuli tap into this representation. Moreover, we assume that the

[13] As in Ramus et al. (1999), we performed each regression once again after exchanging the training and the test sets of sentences, to avoid any asymmetry between the two sets. Scores reported are the average of the two.

Table 7
Language discrimination simulations based on $\%V$ and ΔV.

	Eng.	Dut.	Fre.	Ita.	Cat.	Spa.	Jap.
Dut.	75						
Fre.	*80*	*65*					
Ita.	*85**	*62.5*	45				
Cat.	***95****	*77.5*	52.5	57.5			
Spa.	***77.5***	*75*	55	70	**57.5**		
Jap.	***100*****	*100***	*87.5*	*90**	*80*	*95**	
Pol.	***100*****	*90**	*75*	*80**	***57.5***	***55***	*100***

Scores in boldface are those for which empirical data are available. Cases in italics reflect between-class language pairs.
* For those pairs of languages, one of the two regressions did not converge, i.e. the regression's solution was not unique. This happens when the predictor variables allow to totally separate the two languages (100% correct classifications in the training phase). In this case, we report only the score of the regression that converged.
** For those pairs, none of the two regressions converged. We thus report a 100% score.

Table 8
Language discrimination simulations based on $\%V$, ΔV and ΔC.

	Eng.	Dut.	Fre.	Ita.	Cat.	Spa.	Jap.
Dut.	57.5						
Fre.	*85**	*70*					
Ita.	*75**	*60*	55				
Cat.	***90****	*80**	57.5	57.5			
Spa.	***77.5***	*75*	40	60	**47.5**		
Jap.	***100*****	*100***	*92.5*	*90**	*80*	*95**	
Pol.	***100*****	*90**	*77.5*	*80**	***60***	***60***	*100***

Scores in boldface are those for which empirical data are available. Cases in italics reflect between-class language pairs.
* For those pairs of languages, one of the two regressions did not converge, i.e. the regression's solution was not unique. This happens when the predictor variables allow to totally separate the two languages (100% correct classifications in the training phase). In this case, we report only the score of the regression that converged.
** For those pairs, none of the two regressions converged. We thus report a 100% score.

clusters formed by languages in this space are the basis of the intuitive rhythm types.

General discussion

Based upon their intuitions, linguists have postulated that rhythmical properties of languages are not arbitrary, but may be sorted into a few classes, namely the stress-timed, syllable-timed and mora-timed languages. Ramus et al. (1999) have provided empirical evidence for this notion, showing that variables derived from acoustic/phonetic measurements in the speech signal can account for these rhythm classes. Here, we have tried and provided another type of evidence regarding speech rhythm, i.e., the ability of listeners to discriminate between different types of rhythm.

Ramus et al. (1999) have proposed a model of speech rhythm perception based on the analysis of the proportion of vocalic intervals $\%V$, and this model made precise predictions as to which languages might be discriminated on the basis of their rhythm. It predicted that two languages could be discriminated if and only if they belonged to two different rhythm classes, as traditionally defined.

Nespor (1990) has challenged the classification of two languages, Polish and Catalan, as respectively stress- and syllable-timed; from the measurements of Ramus et al. (1999), it was apparent that the model was likely to accommodate (with variable ΔV) the dissent of Polish, but not that of Catalan. Comparing these two languages with more consensual stress- and syllable-timed languages therefore provided a good empirical test of the model. Here, we have tested the discrimination between Catalan and Polish on the one hand, and reference languages (English and Spanish) on the other. Our use of a speech resynthesis technique (Ramus & Mehler, 1999) ensured that only rhythmical cues were available to subjects. We found that Polish rhythm can indeed be discriminated both from that of English and that of Spanish, but that Catalan rhythm could only be discriminated from that of English and Polish.

Our results suggest that Catalan rhythm is the same as Spanish rhythm, i.e., that Catalan is syllable-timed. Polish, however, seems to be neither syllable- nor stress-timed. New simulations, based on the three variables defined by Ramus et al. (1999), are not only consistent with the present data, but also make a wider range of predictions. It is notably predicted that Polish should be discriminated from all the other languages present in our corpus.

Hence, we view our contribution as having (a) provided empirical evidence for the model proposed by Ramus et al. (1999), (b) successfully adjusted the model to the new data and provided a new set of predictions, (c) provided empirical evidence regarding the typological status of Catalan and Polish. In particular, we suggest that Polish might belong to a rhythm class distinct from the three previously established, and yet to be defined and studied. What might be the phonological characteristics of this new rhythm class? According to Nespor (1990), Polish characteristically has a complex syllable structure, like stress-timed languages, but no vowel reduction, unlike them. This is captured in our model by the newly introduced ΔV variable, which gives a measure of the variability of vowel length. This finding gives credit to Bertinetto (1981), Dasher and Bolinger (1982) and Dauer (1983), who first proposed that vowel reduction played an integral part in the definition of rhythm.

With only eight languages studied so far, the question still remains open whether rhythmic properties shall be described by a discrete set of classes, or by continuous variables. While the present results are adequately described by appealing to four distinct classes, studying more languages might blur the picture. At least, it is now clear that the question can be investigated empirically, by incorporating more languages into the acoustic/phonetic statistics of Ramus et al. (1999), and by studying how listeners classify the rhythm of those languages.

As we have mentioned earlier, speech rhythm is thought to be an important cue to help infants bootstrap the first steps of language acquisition. However, the present work does not allow to draw direct conclusions concerning language acquisition. It is first necessary to study rhythm perception by infants to see whether they are sensitive to the same cues as adult listeners. Relevant data have already begun to accumulate (Mehler et al., 1988; Nazzi et al., 1998; Ramus, Hauser, Miller, Morris, & Mehler, 2000), and more work is in progress (Ramus, submitted). Future studies may investigate for instance whether Polish reflects a particular language rhythm for infants too. Establishing the pattern of language discriminations performed by infants is indeed crucial to understanding precisely how rhythm can be a relevant ingredient of language acquisition.

References

Abercrombie, D. (1967). *Elements of general phonetics.* Chicago: Aldine.

Bertinetto, P. M. (1981). *Strutture prosodiche dell' italiano. Accento, quantit, sillaba, giuntura, fondamenti metrici.* Firenze: Accademia della Crusca.

Bolinger, D. (1965). Pitch accent and sentence rhythm. In *Forms of English: Accent, morpheme, order.* Cambridge, MA: Harvard University Press.

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, *403*, 428-430.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua*, *92*, 81-104.

Dasher, R., & Bolinger, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association*, *12*, 58-69.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51-62.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). The MBROLA Project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *ICSLP'96.* Philadelphia.

Jusczyk, P. W. (1998). Dividing and conquering linguistic input. In M. C. Gruber, K. Olson, & T. Wysocki (Eds.), *CLS 34, vol II: The Panels.* Chicago: University of Chicago.

Ladefoged, P. (1975). *A course in phonetics.* New York: Harcourt Brace Jovanovich.

Lloyd James, A. (1940). *Speech signals in telephony.* London.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (p. 101-116). Mahwah, NJ: Lawrence Erlbaum Associates.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143-178.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, *35*, 666-688.

Nazzi, T. (1997). *Du rythme dans l'acquisition et le traitement de la parole.* Unpublished doctoral dissertation, Ecole des Hautes Etudes en Sciences Sociales.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756-766.

Nespor, M. (1990). On the rhythm parameter in phonology. In I. M. Roca (Ed.), *Logical issues in language acquisition* (p. 157-175). Dordrecht: Foris.

Pallier, C., Dupoux, E., & Jeannin, X. (1997). EXPE: An expandable programming language for on-line psychological experiments. *Behavior Research Methods, Instruments, & Computers*, *29*(3), 322-327.

Pike, K. L. (1945). *The intonation of American English.* Ann Arbor, Michigan: University of Michigan Press.

Ramus, F. (submitted). Perception of linguistic rhythm by newborn infants.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, *288*(5464), 349-351.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America*, *105*(1), 512-521.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265-292.

Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (Ed.), *Linguistic controversies.* London: Edward Arnold.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34-50.