User awareness of *sui generis* pause patterns in simultaneous interpreting

Pradas Macías, E. Macarena
Universidad de Granada, Spain
epradas@ugr.es

*Previous research on fluency suggests that interpreters use a *sui generis* pause pattern in simultaneous interpreting. To identify such a pattern and to explore the users’ awareness of it, we carried out two studies. The first one is an observational study aimed at finding out which lexical categories in a source text (ST) are more likely to prompt interpreters to insert pauses in the target text (TT). In our sample, pauses appeared more frequently and more consistently following the rendering of nouns. The second study is an experiment seeking to determine whether deviations from this pattern could be detected by a group of listeners and whether they influenced their evaluation of interpreting quality. When the pauses following ST-noun phrases were removed from the TT, the participants described the pause pattern as unexpected. In addition, they characterized certain paralinguistic quality criteria as unexpected but rated them more favourably than in the original version of the TT.*

[A] Introduction

The importance of interpreting quality and the difficulty to describe its composition was pointed out already more than 30 years ago (cf. AIIC 1979). Ever since, quality has become one of the most cohesive research fields in conference interpreting (Gile 2000: 305). A large number of studies on quality employ surveys revolving around a more or less stable catalogue of criteria introduced by Bühler (1986). However, it has often been claimed that these criteria are ill-defined and probably overlap each other (e.g. Garzone 2003; Collados Aís et al. 2011) in particular those related to both form and content, like fluency of delivery. This awareness led to the adoption of a top-down approach aimed at defining fluency in terms of subcriteria (Pradas Macías 2003b).

An exhaustive review of the literature on fluency (Pradas Macías 2003a) showed that (silent) pauses appear consistently in the various definitions of this quality criterion. In psychology and psycholinguistics, pauses play a role as an expression of certain pathologies. In language-related fields, they have been studied to gain insight into production and perception processes (cf. Barik 1968, 1972, 1973; Darò 1995).
In simultaneous interpreting, there is a broad consensus among researchers that pauses are one of the most salient features of fluency (e.g. Kopczyński 1981; Pöchhacker 1994, 1995; Rennert 2010). In some studies, pauses have been spontaneously mentioned by users when asked to define fluency (Pradas Macías 2003b, 2006a). In this field, pauses have been studied from a range of different perspectives (Pradas Macías 2006a):

a) The effect on the interpreter of the pauses occurring in the source speech (e.g. Barik 1972, 1973, 1975; Tissi 2000).

b) The use of source-speech pauses by the interpreter (e.g. Alexieva 1988; Shlesinger 1994; Mead 2000).

c) The impact of silent pauses on the users’ assessment of interpreting quality (Ahrens 1998), and more specifically of the fluency of the interpreter’s delivery (Pradas Macías 2003b).

One obvious research question relates to the motivations behind the interpreters’ pause insertion and whether there are similarities between the pause patterns of different interpreters when interpreting one and the same source speech (ST). With this in mind, we examined the interpretations into Spanish (TT) of a speech given in German carried out by four professional interpreters of two different backgrounds, the European Commission and the private market (Pradas Macías 2006b). Generally speaking, the pause patterns used by these interpreters seemed to follow the same basic trend. In a later study (Pradas Macías 2009) we tried to reproduce this finding in the interpretations of several speeches of the same language combination (German into Spanish). The results suggest that interpreters might use a sui generis pause pattern in simultaneous mode, and that it does not seem to match the pause pattern of the ST.

This lack of correlation between both speeches begs the question whether the pause pattern in the TT might be induced by the linguistic features of the ST rather than by its paralinguistic features, in particular by the lexical categories of the words adjacent to the pauses. This motivated Study A described below, and Study B was carried out to determine whether users are aware of such a pattern – more specifically, if they perceive an altered pause pattern as unexpected, and whether this manipulation also affects their evaluation and their perception of standard quality criteria measured against their expectations.

[A] Hypotheses and objectives

For the purposes of this study, let us define pauses as an interruption in the flow of speech regardless of cause or intention (breathing, hesitation, rhetorical effect etc.). The research questions outlined before find their concrete expression in the following hypotheses:

1. Existence of pause-triggering lexical categories: The interpretations of the lexical items in the ST are more frequently adjacent to pauses when the items belong to certain lexical categories.

2. Perception of pause-pattern deviance: When pauses adjacent to the interpretation of lexical items from a pause-triggering category are deleted from the audio recording of a TT, users tend to perceive the realization of the pause pattern (and of standard quality criteria) as unexpected.
3. Impact on quality evaluation: The pause manipulation leads to changes in the users’ ratings of the pause pattern (and of standard quality criteria).

We tested the first hypothesis in an observational study (Study A) which involved the following objectives:

- Selection of a sample of audio-recordings of STs and their respective TTs
- Identification of all lexical items in the TTs that are adjacent to pauses
- Identification of the lexical items in the STs that can be mapped to the lexical items identified in the TTs
- Identification of the lexical categories of the items identified in the STs
- Counting of the occurrences of lexical items in each category

The second and the third hypotheses were tested in an experiment (Study B), with the following objectives:

- Selection of one of the speeches of the sample of Study A
- Selection of the most prominent pause triggering category found in the speeches of Study A
- Identification of the lexical items belonging to the selected category in the ST of the selected speech
- Creation of a manipulated version of the TT of the selected speech by means of deletion of pauses adjacent to those lexical items
- Drafting of a questionnaire addressing expectation matching and the evaluation of the pause pattern and of several quality criteria
- Selection of experimental subjects
- Testing of the subjects’ perception of the manipulated TT as opposed to the control TT, in terms of expectation matching and of quality evaluation.

[A] Study A: Existence of pause-triggering lexical categories

[B] Material

The sample used in Study A consists of the recordings and the transcriptions of five speeches taken from the ECIS 2003 corpus (ECIS Research Group 2003). They were held in German and interpreted into Spanish during a plenary session of the European Parliament, and we selected them because they were held on the same day, the speech rate of their respective TTs was similar and they dealt with the same subject, namely the Policy Strategy of the European Commission 2004 and Budget Guidelines.

The respective speech rates are shown in Table 1. To control for differences in pause length, we ensured that the TTs had a similar speech rate (mean value 124.42 words/min, standard deviation 4.05).

---

1 A corpus was compiled by the ECIS Research Group, based at the University of Granada, consisting of the recordings of the speeches given in German, French and English during three plenary sessions of the European Parliament (10th–13th March 2003), and their respective simultaneous interpretations into Spanish.
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ST</strong></td>
<td>159.65</td>
<td>130.28</td>
<td>127.00</td>
<td>186.54</td>
<td>112.92</td>
</tr>
<tr>
<td></td>
<td>404</td>
<td>525</td>
<td>135</td>
<td>220</td>
<td>144</td>
</tr>
<tr>
<td><strong>TT</strong></td>
<td>126.39</td>
<td>125.18</td>
<td>117.76</td>
<td>128.53</td>
<td>124.25</td>
</tr>
<tr>
<td></td>
<td>851</td>
<td>1185</td>
<td>263</td>
<td>437</td>
<td>292</td>
</tr>
<tr>
<td></td>
<td>404</td>
<td>568</td>
<td>134</td>
<td>204</td>
<td>141</td>
</tr>
</tbody>
</table>

Table 1: Speech rates and duration of the speeches in the sample

[B] Method

The pauses were measured with Adobe Audition 1.5. Since the pause distributions in each of the five TTs do not correlate with the distributions in the respective STs, the ST pause pattern may be ruled out as a confounding variable.

We discarded all pause occurrences shorter than 250 ms, following the rationale set out in former research (e.g. Tissi 2000: 105). In addition, we selected a given pause only if it was adjacent to a lexical item in a TT that could be unequivocally identified as a verbatim interpretation of some item in the ST. As shown in Table 2, this had only a slight effect on the samples, as a minimum of 85% of the pauses fulfilled this requirement in every one of the five TTs.

<table>
<thead>
<tr>
<th></th>
<th>TT1</th>
<th>TT2</th>
<th>TT3</th>
<th>TT4</th>
<th>TT5</th>
</tr>
</thead>
<tbody>
<tr>
<td>pauses</td>
<td>108</td>
<td>149</td>
<td>41</td>
<td>66</td>
<td>54</td>
</tr>
<tr>
<td>pauses / words (%)</td>
<td>12.69</td>
<td>12.57</td>
<td>15.59</td>
<td>15.10</td>
<td>18.49</td>
</tr>
<tr>
<td>pauses selected</td>
<td>101</td>
<td>146</td>
<td>37</td>
<td>58</td>
<td>46</td>
</tr>
<tr>
<td>pauses selected (%)</td>
<td>93.52</td>
<td>97.99</td>
<td>90.24</td>
<td>87.88</td>
<td>85.19</td>
</tr>
<tr>
<td>pauses selected / words (%)</td>
<td>11.87</td>
<td>12.32</td>
<td>14.07</td>
<td>13.27</td>
<td>15.75</td>
</tr>
</tbody>
</table>

Table 2: Pauses (≥ 250 ms) adjacent to lexical items identified as verbatim interpretations

[B] Results

For each of the pauses and lexical items selected in the TT, the corresponding item was located in the ST. After identifying the lexical category for every one of these items, the occurrences were counted in each category. *Conditional conjunctions* are categorised separately from *conjunctions*, and *temporal adverbs* separately from *adverbs* because they occur at different positions within the sentences. The frequencies are shown as relative percentages in Table 3.

---

2 This legacy version is especially convenient for single-track edition. The Adobe Audition series is offered by Adobe Labs, http://www.adobe.com/.
Table 3: Relative frequencies (%) of ST items in each lexical category interpreted as TT items appearing before/after pauses

<table>
<thead>
<tr>
<th></th>
<th>ST1 before</th>
<th>ST1 after</th>
<th>ST2 before</th>
<th>ST2 after</th>
<th>ST3 before</th>
<th>ST3 after</th>
<th>ST4 before</th>
<th>ST4 after</th>
<th>ST5 before</th>
<th>ST5 after</th>
</tr>
</thead>
<tbody>
<tr>
<td>nouns</td>
<td>35.6</td>
<td>5.9</td>
<td>35.4</td>
<td>7.4</td>
<td>43.2</td>
<td>2.7</td>
<td>36.2</td>
<td>10.3</td>
<td>47.8</td>
<td>4.4</td>
</tr>
<tr>
<td>verbs</td>
<td>27.7</td>
<td>16.8</td>
<td>34.7</td>
<td>16.3</td>
<td>13.6</td>
<td>16.2</td>
<td>38.0</td>
<td>10.3</td>
<td>26.0</td>
<td>10.8</td>
</tr>
<tr>
<td>articles</td>
<td>2</td>
<td>9.9</td>
<td>2</td>
<td>12.2</td>
<td>0</td>
<td>16.2</td>
<td>0</td>
<td>8.6</td>
<td>0</td>
<td>21.7</td>
</tr>
<tr>
<td>prepositions</td>
<td>2</td>
<td>6.9</td>
<td>4.8</td>
<td>13.6</td>
<td>2.7</td>
<td>10.8</td>
<td>0</td>
<td>6.9</td>
<td>2.2</td>
<td>19.6</td>
</tr>
<tr>
<td>conjunctions</td>
<td>9.9</td>
<td>22.8</td>
<td>4.8</td>
<td>15.0</td>
<td>5.4</td>
<td>27.1</td>
<td>5.2</td>
<td>24.1</td>
<td>4.4</td>
<td>10.8</td>
</tr>
<tr>
<td>pronouns</td>
<td>10.9</td>
<td>16.8</td>
<td>1.4</td>
<td>13.6</td>
<td>5.4</td>
<td>5.4</td>
<td>5.2</td>
<td>6.9</td>
<td>2.2</td>
<td>8.7</td>
</tr>
<tr>
<td>relative pron.</td>
<td>0</td>
<td>4.6</td>
<td>0</td>
<td>4.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>adjectives</td>
<td>3</td>
<td>5.9</td>
<td>2</td>
<td>9.5</td>
<td>5.4</td>
<td>10.8</td>
<td>10.3</td>
<td>19.0</td>
<td>13.0</td>
<td>10.8</td>
</tr>
<tr>
<td>adverbs</td>
<td>5.9</td>
<td>8.9</td>
<td>4.8</td>
<td>4.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>proper nouns</td>
<td>3</td>
<td>0</td>
<td>2.7</td>
<td>0.7</td>
<td>16.2</td>
<td>2.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>negative part.</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.7</td>
<td>1.7</td>
<td>1.7</td>
<td>0</td>
<td>2.2</td>
</tr>
<tr>
<td>modal part.</td>
<td>0</td>
<td>0</td>
<td>1.4</td>
<td>1.4</td>
<td>2.7</td>
<td>2.7</td>
<td>1.7</td>
<td>1.7</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>condit. conj.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.7</td>
<td>0</td>
<td>1.7</td>
<td>3.5</td>
<td>0</td>
<td>2.2</td>
</tr>
<tr>
<td>temporal adv.</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
<td>0</td>
<td>0</td>
<td>2.7</td>
<td>0</td>
<td>3.5</td>
<td>0</td>
<td>2.2</td>
</tr>
<tr>
<td>interrog. part.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>numbers</td>
<td>0</td>
<td>0</td>
<td>5.4</td>
<td>0</td>
<td>2.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.2</td>
<td>4.4</td>
</tr>
</tbody>
</table>

[B] Discussion

In the sample, the most prominent pause triggering effect can be found in three lexical categories, namely in the nouns and the verbs located “before” a pause in the TT and in the conjunctions occurring “after” a pause. In terms of interpreting strategies, one possible explanation for this is that the interpreter waited after completing an information unit or before starting to interpret a new one.

There are some interesting correlations between TT duration and pause occurrences. First, the most salient result in Table 3 is that nouns are located frequently “before” pauses, but this is even more so in the case of the shortest speeches, TT3 and TT5. Since these two have a proportionally larger number of pauses ≥ 250 ms (see Table 2), this result is consistent with the fact that the presence of pauses allows the interpreter a longer ear-voice span (EVS) after interpreting a phrase. Second, pronouns appear more often “after” pauses in the longest speeches TT1 and TT2. Third, articles and adjectives in the ST occur more frequently “after” a pause in the shortest TTs, which suggests that the interpreter was waiting for the next information unit, given that in German attributive adjectives typically precede their nouns. These pauses may well be regarded as “pauses for thought” or “for utterance planning” (Ahrens 1998: 224). Fourth, no occurrences of relative pronouns or of adverbs could be identified in the neighbourhood of the pauses selected in the two shortest TTs.
Finally, a close examination of the speech transcripts shows that certain lexical categories with low frequencies are followed by pauses which frequently give way to a notably wide variety of lexical categories. In particular, in the rare occasions where conjunctions, articles and adjectives appear “before” a pause, the category of the subsequent lexical item is hard to predict. Considering the high level of expertise of the interpreters involved, this phenomenon can probably be attributed to text-specific factors. Some of the pauses may be due to the lengthening of the EVS as an anticipation strategy or to difficulties in predicting the information ahead, as found in former research on simultaneous interpretation from Finnish to Swedish:

The interpreter makes unfilled pauses for the sake of the listener as well as for planning and monitoring his own output. However, there is yet another dimension: the interpreter may pause in order to wait for the speaker to speak so that he would have something to interpret. This may happen when the interpreter has followed very close to the speaker who then starts an utterance the outcome of which is not immediately clear. (Ovaska 1987: 81)

In the two shortest speeches, TT3 and TT5, articles appear more often after a pause, and there are many more noun occurrences before than after pauses. This indicates that the interpreters may have resorted to anticipation more frequently, possibly due to predictable redundancies:

A highly redundant text, such as an oral communication, is predictable. Hence, the psychological machinery of probability prediction entails some kind of correlation between the probability of certain linguistic (phonetic, prosodic, verbal and syntactic) or semantic (meaning and sense) developments in the utterance and coherent discourse at the stage of perception (communication between the speaker and the interpreter), and the anticipatory synthesis of the TL (target language) message. (Chernov 1994: 145)

Finally, it is only in TT4 that verbs appear more often before pauses than nouns. In this speech, there are no articles adjacent to pauses and adjectives tend to occur after pauses.

[A] Study B: Perception of pause-pattern deviance and impact on quality evaluation
[B] Material

All of the five TTs of Study A had been previously evaluated with regard to a number of quality criteria by two researchers of the University of Granada who collaborated with our research group by evaluating the speeches and their interpretations forming part of the ECIS 2003 corpus. They had been requested to mention the most salient positive and negative features of each interpretation and to fill in a questionnaire based on a Likert-type scale (from 1 worst to 5 best) to assess the following quality criteria: style, cohesion and fluency, accent and diction, voice, grammar, intonation, fidelity, and terminology. In the case of TT5, fluency was mentioned by both judges as the most salient of the favourably rated features, and both fluency and intonation received a rating of 4 out of 5 points.

In addition, TT5, one of the two shortest speeches (both in terms of number of words and of duration) contained the largest number of pauses equalling or exceeding 250 ms. This was also true after discarding some of the pauses, with the proportion between number of pauses and number of words decreasing from 18.49 to 15.75%
(see Table 2). As a result, pause removal was likely to cause a more noticeable effect on this TT than on the others.

For these reasons, we decided that Study B would focus on TT5 to check the users’ expectations with regard to pause patterns and other speech features. The experiment would probe their perceptions of a manipulated version of TT5 and of TT5 itself, acting as a control version. The manipulation consisted in deleting the pauses preceded by interpretations of nouns, with one exception: the pauses located at clause boundaries were exempted from deletion because this might have introduced a confounding variable. Moreover, the deletion was not absolute: we reduced each pause to a bare minimum (a silent gap of 20 ms), so as to allow the interpreter’s articulation to still sound natural. The changes were made using Adobe Audition.

[B] Subjects
In an unpublished study developed during the academic years 2008/2009 and 2009/2010, 18 students enrolled in a one-year Masters’ in translation and interpreting research at the University of Granada were asked to rate the quality of an interpreted speech. The results suggest that interpreting students are quite exacting when evaluating an interpreter’s performance as for both content and form.

With this in mind, the participants for Study B were recruited among the 17 students enrolled in a one-year Master’s in conference interpreting at the University of Granada, during their second semester. 16 of them were available, but two of them were excluded from the sample after the experiment, one due to missing data and one because he selected two response options when only one was expected.

The final sample included 14 subjects. Six of them listened to the original TT5 (control group) and eight to the manipulated version (treatment group). 7 of the students were male and 7 female, 11 were aged 22-28 and 3 were aged 31-40 years.

[B] Questionnaire

After listening to the recording, we asked each participant to complete a questionnaire consisting of 10 sections.

Eight sections were devoted to a selection of quality criteria: overall quality, accent, fidelity, intonation, logical cohesion, fluency, voice quality and (phonological) diction. The results of previous experiments (Pradas Macías 2003, 2007) advised to exclude style and grammar, as they seemed to be unaffected by pause pattern alterations. We also discarded terminology because the selected speech is of a rather non-technical nature, and fidelity is understood as a broad term encompassing completeness (Nobs et al. 2011).

In each of these sections, the subjects were asked to rate the respective criterion on a 5-point rating scale. Moreover, they were asked an open-ended question

---

3 This was a replication of a study by Pradas Macías (2003b), in which a panel of experts where requested to assess the quality of three recordings of a simultaneous interpretation that differed only with regard to the frequency of silent pauses.
on whether it had matched their expectations. This question had been tested for intelligibility by a senior researcher in the run-up of the experiment.

In another section, the subjects were asked to rate the *pauses* on a 5-point rating scale in terms of frequency (from too few to too many), of length (from too short to too long) and of position (from inappropriate to appropriate). Note that, while the scale for position was increasing, the scales for frequency and length were symmetric, with 3 as their maximum value. This section also included a question asking the participants whether they perceived the pause pattern as expected or as unexpected.

A final section included a group of questions addressing different aspects regarding the interpreter’s professionalism. A detailed account of their answers, already announced by Pradas Macías and García Becerra (2013), will be the subject of a separate paper by the same authors (in prep.).

[B] Results

[C] Evaluation of quality criteria

![Figure 1: Quality criteria ratings given by the control group](image)

Since the quality criteria were rated on an ordinal scale, the ratings can be interpreted conveniently with the help of a reverse cumulative graph, in which the frequencies correspond to the number of subjects whose ratings equal or exceed a given value. The subjects’ evaluation of the control version (Figure 1) was especially favourable with respect to *accent* and *logical cohesion*, while *intonation* ranked lowest, with all ratings equal or lower than 3. The control group gave no 1-rating in any category, and *overall quality, accent* and *diction* were rated 3 or higher by all respondents in this group.
The manipulated version (Figure 2) received the best ratings for *accent* as well. *Intonation* scored lowest and was also the only criterion receiving a 1-rating. The criterion was followed by diction. Like *accent*, the criteria *fidelity* and *fluency* received very high ratings, all of them equalling or exceeding 4. *Overall quality* and *logical cohesion* were always rated 3 or higher, albeit seldom reaching the top rating.

Overall, the manipulated version was rated higher in most categories. Two salient exceptions are *overall quality*, behaving similarly, and *diction*, perceived as better in the control version. The effect of pause deletion is especially apparent in the case of *fidelity* and *fluency*, and inconsistent for *logical cohesion* and *intonation*.

[C] Evaluation of pauses

Let it be reminded that, by question design, pause *frequency* and *length* were rated on a symmetric scale (from 1, “too few pauses” to 5, “too many pauses”, with 3 as the best value), and *position* on an increasing scale (from 1 “inappropriate” to 5 “appropriate”). The subjects evaluating the control version (Figure 3) often judged *position* neutrally, and tended to give slightly negative ratings as for the other two features: as many as half of them considered that there were too few pauses or too short ones.
As for the manipulated version (Figure 4), the rating distribution with regard to position was slightly right-tailed and more even than in the control. This time, the frequency of pauses was deemed adequate by half of the respondents, but the ratings were still skewed to the left for both frequency and length. In summary, the deletion of pauses seems to have prompted a more positive evaluation, especially with regard to frequency.

[C] Expectation matching

We construed the participants’ answers to the question whether a given criterion or the pause pattern had matched their expectations, coded them into categories and counted the occurrences in each category. The frequency counts are shown in Tables 4 and 5. We used the following category codes: Yes (expectation has been matched), No (not matched), Yes* (the subject’s statement was affirmative, but it did not include the word “yes”), No* (negative statement without an explicit “no”), Unclear (ambiguous statement) and No answer.

<table>
<thead>
<tr>
<th>Category</th>
<th>Yes</th>
<th>Yes*</th>
<th>No</th>
<th>No*</th>
<th>Unclear</th>
<th>No answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>overall quality</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>accent</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>fidelity</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>intonation</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>logical cohesion</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>fluency</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>voice</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>diction</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>pauses</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: Expectation matching for the control group (6 subjects), in absolute frequencies

At least half of the subjects rating the control version felt that their expectations had been matched in all categories except for voice, which elicited ambiguous answers from the respondents. Even if the No* answers were counted as No answers, a 50% non-matching rate would be reached only in the diction category. The highest expectation-matching rate can be found for accent, followed by overall quality, fidelity, logical cohesion and fluency.
Table 5: Expectation matching for the treatment group (8 subjects), in absolute frequencies

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>Yes*</th>
<th>No</th>
<th>No*</th>
<th>Unclear</th>
<th>No answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>overall quality</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>accent</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>fidelity</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>intonation</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>logical cohesion</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>fluency</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>voice</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>diction</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>pauses</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

As for the manipulated version, the expectations of less than 50% of the participants were matched with respect to intonation, voice, diction and pauses. The results for voice are difficult to interpret because three out of eight subjects did not answer the question at all. The situation is similar in the case of diction, as two of them did not respond and the opinions of the other two are unclear. Nevertheless, this is the only criterion exhibiting a clearly negative leaning (1 positive, 3 negative rating). The highest expectation-matching rates can be found for overall quality, accent, fidelity and accent, while the values for pauses are polarized.

In comparison to the control, the deletion of pauses in the manipulated version seems to have improved expectation matching especially with respect to fidelity, to logical cohesion and fluency, and to a lesser extent to intonation and pauses.

[C] Do evaluation and expectation-matching correlate?

For virtually all criteria, good ratings correlate with the confirmation of expectation matching. However, with regard to intonation and pauses, the manipulated version was rated more favourably than the control version, but the realization of those features tended to be perceived as unexpected.

[B] Discussion

The lowest ratings of intonation and of voice referred to the control version. Interestingly, its original version TT5 had not received any negative ratings with regard to either of both criteria when it was evaluated by two experts for the ECIS 2003 corpus. This result leads us to think that our subjects may probably be especially sensitive to these criteria as they were receiving specialized training as interpreting students at the time of the experiment. In addition, both criteria are closely related with one another (cf. Collados Aís et al. 2011).

By contrast, in the manipulated version, both of these criteria were rated more favourably, just like most of the other criteria (except for diction). This result is in agreement with the third hypothesis, which states that any alterations to the pause pattern may have an impact on quality evaluation.
Moreover, these findings are consistent with the outcome of a former study (Pradas Macías 2003) in the sense that pause manipulation bears heavily upon intonation. With this in mind, the positive influence on fidelity is not as surprising as it may seem: if altering the pause pattern improves not only the subjects’ opinion about the interpreter’s intonational pattern, but their actual perception of it, comprehension might also be enhanced. This is consistent with Shlesinger’s (1994: 232) conjecture, which states that form may affect the evaluation of content and was first supported by evidence in an experiment on intonation and fidelity (Collados Aís 1998). This possibility also echoes other findings showing that pause patterns commonly used by professional interpreters do not impinge on the evaluation of fluency (Pradas Macías 2003). In fact, in this experiment, pause removal was accompanied by a slight improvement of the subjects’ ratings of fluency. This is a salient result given that this was the criterion that received the best rating in the control version. On its part, the fact that the rating of overall quality was similar for both versions does not come as a surprise, as TT5 had been selected for Study B from other speeches precisely because of its high quality.

The pauses also received better ratings in the manipulated version than in the control version, especially with regard to their frequency (Figure 4). However, since the scale used for frequency and length was symmetrical, these features might be more susceptible to a central bias than position, rated on an ordinal scale. Moreover, manipulation seemed to have a levelling effect on the rating distribution for position, effectively improving its rating.

The second hypothesis could be confirmed as well, as the changes made to the pause pattern led the subjects to perceive the realization of various speech features as unexpected (Tables 4 and 5). With regard to the pause pattern itself, the participants reported that their expectations had been matched more often in the manipulated version than in the control. The outcome is similar for intonation, which received a higher rating when pauses had been removed, but was perceived as unexpected more often than in the control version. These results suggest that the participants were aware of the existence of a sui generis pause pattern in simultaneous interpreting, albeit the interpreter’s delivery may receive a better rating when it deviates from this pattern. This supports the conjecture posed by Butcher (1980: 85) according to which intonation is influenced by syntactic structure and pause patterns.

Therefore, it is plausible that removing pauses that cannot be interpreted as indicators of a long ear-voice-span may prompt a negative evaluation of pause patterns and of the interpreter’s delivery as a whole. Considering that a sui generis pattern has been identified in favourably rated interpretations carried out by professional practitioners (Pradas Macías 2009), the subjects might be punishing the pattern deviations that cannot be explained in terms of their beliefs about the interpreting process:

The results stress the existence of, and a subject’s adherence to, an implicit temporal rule in conversation, which states that an excessive gap will be perceived as deviant [...] .(D’Urso and Zammuner 1990: 42)

[A] Conclusions
The results suggest that listeners are aware of the existence of a sui generis pause pattern in simultaneous interpreting. Moreover, the pause removal carried out in the
experiment had a slight positive impact on quality evaluation. This result may appear to be paradoxical, but it also supports the notion that the line between linguistic and paralinguistic aspects of fluency is blurred.

The outcome of Study A suggests that certain lexical categories act as pause triggers: the simultaneous interpretations of nouns, verbs and conjunctions in the ST are more frequently adjacent to pauses in the TT than those belonging to other categories. This might be due to anticipation and décalage strategies on the part of the interpreter, an idea supported by the fact that after the pauses following nouns and verbs, the occurrence of conjunctions, articles and adjectives was more likely. On their part, pronouns, proper nouns and numbers appeared less frequently around pauses. When they did occur next to a pause, they were usually placed after it, possibly because they complement the information borne by nouns and other parts of speech.

However, it must be observed that in this research, the explanations for the presence of pauses have been searched for within a short range, although the causes for a pause occurrence may well extend beyond the scope of a clause or an utterance, as suggested by Baars (1980: 325), who argued that the temporal features of oral production are conditioned by forward and backward interactions of both proximal and distal nature.

In conference interpreting, it is reasonable to expect that most subjects will become familiar with the sui generis pause pattern of interpreters at some point. In the view of this, pause manipulation (in particular, deletion and insertion) may well affect the perception of interpreting quality. The method used in this study could be refined in two ways. On the one hand, by using a more detailed catalogue of lexical categories, defined in terms of their syntactical and even semantic functions. On the other hand, by automating the whole process. First, pauses could be detected automatically using a digital audio editor, exporting their positions and lengths to a so-called cue list, as done in former research (Pradas Macías 2012). Second, if signal-noise ratio and the speaker’s language and accent allow, speech transcription could be automated by means of an automatic phonetic alignment tool such as P2FA (Yuan & Liberman 2008). Third, the resulting time-aligned text would be imported into a lemmatizing software to identify the lexical categories adjacent to each pause. Fourth, the pauses to be manipulated would be identified. Finally, these pauses would be targeted with the help of a custom script programmed as an add-in for an acoustic analysis application such as Praat. Under optimal circumstances, these steps would save researchers valuable time by reducing manual intervention to transcript proofreading and pause selection.

Moreover, since the pauses were actually not removed but replaced with 20-ms silent gaps, the use of fixed thresholds to define pauses may be put in question. It is the researcher’s contention that pause detection is based rather on function than on length. The fact that, for these two studies, only pauses over 250 ms were considered, might be the reason why no pauses could be found around certain lexical categories. Indeed, a close examination of TT5 demonstrates that there were indeed shorter pauses adjacent to categories such as relative pronouns and adverbs, which were especially affected by using that threshold.

Finally, different pause lengths could be tested in future experiments to find a threshold of influence on the perception of intonational and syntactical patterns. It is quite surprising that intonation and voice received slightly higher ratings in both the
control and the manipulated version of Study B. This hints that it might also be appropriate to complement the analysis of pause patterns with the examination of other acoustic features of speech.

[A] Acknowledgments

Research described in this paper was supported by the ECIS Research Group (Evaluación de la calidad en interpretación simultánea), ref. HUM 560 of the Andalusia Council of Science.

[A] References


BÜHLER, Hildegund (1986): “Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters”, Multilingua 5 (4), 231–236.


PRADAS MACÍAS, E. Macarena (2012): “Preparing experimental material for the study of pauses: instrumental comparison”, *EST Symposium Same Place, Different Times*, University of Vienna, 27th–28th September 2012.


PRADAS MACÍAS, E. Macarena & GARCÍA BECERRA, Olalla (in prep.): “Perceived professionalism and *sui generis* pause patterns in simultaneous interpreting.”


