

# The limits of *expectations vs. assessment* questionnaire-based surveys on simultaneous interpreting quality: the need for a gestaltic model of perception

GREGORIO DE GREGORIS  
Università di Trieste  
gregorio.degregoris@phd.units.it

## ABSTRACT

This paper deals with a literature review on Interpreting Studies on the evaluation of quality in simultaneous interpreting. The analysis has been made in function of a proposal for a gestaltic evaluation, and the studies have been divided into two categories: quality expectations surveys (ideal expression of preferences) and quality assessment surveys (judgments after a real experience). Conclusions are drawn from the comparison of results of quality expectations and quality assessment surveys, and the interrelation and interdependence of quality criteria. Conclusions lead to the need for a model to elicit a gestaltic perception of simultaneous interpreting quality. Such a model has already been developed and has been applied to a proposal for a new questionnaire for a gestaltic quality assessment of TV broadcast simultaneous interpreting (De Gregoris, submitted).

## KEYWORDS

Interpreting, survey, expectation, assessment, perception.

## INTRODUCTION

In surveys on quality expectations (ideal evaluation) of simultaneous interpretation subjects tend to give more importance to those features that have been labelled as ‘content-related aspects’, like ‘transmission of the original sense’, ‘logical coherence’, ‘terminology’; while in surveys on quality assessment (after a real experience of the phenomenon) subjects still give more importance to the so called ‘content-related aspects’, but in this case they give higher ratings to those feature that have been labelled as ‘form-related aspects’, like ‘fluency’, ‘voice’, ‘rhythm’, which have an impact on the assessment of the ‘original sense’, ‘coherence’, ‘accuracy’ and on the evaluation of the ‘overall quality’ of an interpretation. This is particularly evident when perception is received through an audiovisual medium (Russo 2005), or when parameters are individually manipulated to measure the incidence of manipulation on other parameters (Collados Aís *et al.* 2007). For this reason the need for a gestaltic approach to assess the quality of interpretation has been called upon (Garzone 2003; Soler Caamaño 2006; Iglesias Fernández 2013).

### 1. QUESTIONNAIRE-BASED SURVEYS ON QUALITY EXPECTATIONS IN SIMULTANEOUS INTERPRETING

This section deals with a literature review of studies on quality expectations of simultaneous interpreting (SI), i.e. questionnaire-based survey on the ideal evaluation of interpretation, or the expression of preferences of quality on the basis of pre-definite or spontaneous criteria. These quality criteria are linked to the items of the questionnaires, that require the respondents either to rate each of them on a numeric scale, to rank them in order of importance or to comment on them through open questions. Subjects are interpreters, delegates of EU institutions, other users like medical doctors, engineers, etc.

#### 1.1. BÜHLER (1986)

The first questionnaire on quality expectations in simultaneous interpreting (SI) was built by Hildegund Bühler (1986) and administered to 41 interpreter-members of the *Association internationale des interprètes de conférence* (AIIC) and 6 members of the *Commission des admissions et du classement linguistique* (CACL) of AIIC. Quality criteria were grouped in “linguistic (semantic)” and “extra-linguistic (pragmatic)”. The “linguistic (semantic)” criteria were: *native accent, pleasant voice, fluency of delivery, logical cohesion of utterance, sense consistency with original message, completeness of interpretation, correct grammatical usage, use of correct terminology, use of appropriate style*; while the “extra-linguistic (pragmatic)” criteria were: *thorough preparation of conference documents, endurance, poise, pleasant ap-*

*pearance, reliability, ability to work in a team, positive feedback from delegates.* The study by Bühler (1986) revealed that the “linguistic” criteria received the highest ratings from respondents: *sense consistency with original message* was the first, followed by *logical cohesion of utterance, completeness of interpretation, use of correct terminology, correct grammatical usage, fluency of delivery* (1986: 232). The criterion *reliability* received the highest rating among “extra-linguistic” criteria, followed by *thorough preparation of conference documents and ability to work in a team.* The criteria of *native accent, pleasant voice, use of appropriate style, endurance, poise and pleasant appearance* “were considered desirable in most cases but not essential” (1986: 233).

Items of the questionnaire (Bühler 1986)	highly important	important	less important	irrelevant
1. native accent	23%	47%	28%	
2. pleasant voice	28%	61%	9%	
3. fluency of delivery	49%			
4. logical cohesion of utterance	83%			
5. sense consistency with the original message	96%			
6. completeness of interpretation	47%	49%		
7. correct grammatical usage	48%	50%		
8. use of correct terminology	49%			
9. use of appropriate style	7%	68%	15%	
10. thorough preparation of conference documents	73%			
11. endurance				
12. poise				
13. pleasant appearance				
14. reliability	73%	22%		
15. ability to work in a team	47%	49%		
16. positive feedback from delegates				
<i>other criteria (please specify):</i>				

Table 1. Items of the questionnaire by Bühler (1986) with the respective preferences assigned to each item by respondents (in percentages) in a four-point labelled scale – my calculations based on the data in the paper.

1.2. KURZ (1989; 1993)

After Bühler, Kurz (1989) used a questionnaire-based survey using the first eight “linguistic (semantic)” criteria in Bühler (1986), but the questionnaire was administered to a specific group of users: medical doctors. The same questionnaire some years later (Kurz 1993) was administered to other two different user groups: engineers and delegates of the Council of Europe. The surveys conducted by Kurz (see table 2) revealed that different user groups had different degrees of expectations of quality in simultaneous interpreting, or, in other words, quality criteria varied according to the user group. Interpreters had higher expectations than other groups, because they assigned higher rates to the criteria than other groups. Nonetheless, all user groups, on average, assigned more importance to *sense consistency with original message*, followed by *logical cohesion*, *use of correct terminology* and *completeness of interpretation*; while the last positions of the ranking were occupied by *fluency of delivery*, *correct grammatical usage*, *pleasant voice* and *native accent*.

Criterion (rated on a 4-point scale)	Bühler '86 (in Kurz '93)	Kurz '89 (in Kurz '93)	Kurz '93		
	Interpreters (n=47)	Medical doctors (N=47)	Engineers (N=29)	CEdelegates (N=48)	average
1. native accent	2.9 (8)	2.3 (8)	2.2 (7)	2.08 (8)	2.365 (8)
2. pleasant voice	3.085 (7)	2.6 (6)	2.4 (6)	2.396 (7)	2.6 (6)
3. fluency of delivery	3.468 (4)	2.9 (5)	2.966 (4)	3.208 (5)	3.1 (5)
4. logical cohesion	3.8 (2)	3.6 (1)	3.1 (3)	3.3 (4)	3.458 (2)
5. sense consistency with original message	3.957 (1)	3.6 (1)	3.655 (1)	3.6 (2)	3.69 (1)
6. completeness of interpretation	3.426 (5)	3.0 (4)	2.9 (5)	3.458 (3)	3.2 (4)
7. correct grammatical usage	3.38 (6)	2.4 (7)	2.03 (8)	2.688 (6)	2.6 (6)
8. use of correct terminology	3.489 (3)	3.4 (3)	3.138 (2)	3.729 (1)	3.4 (3)
average	3.44	3.0	2.8	3.06	3.06

Table 2. “Shows the significance attributed to the different criteria by the four groups of subjects” (Kurz 1993: 16). Numbers indicating ranking positions (in brackets) are mine.

### 1.3. KURZ AND PÖCHHACKER (1995)

Kurz and Pöchhacker (1995) used Bühler’s eight “linguistic (semantic)” criteria for a questionnaire-based survey on quality expectations in television interpreting. This time, the questionnaire was administered to “a group of representatives of Austrian and German TV organizations” (1995: 351). In this case, TV professionals’ expectations for quality in “simultaneous interpreting for live television broadcasts” (1995: 350) were higher than conference participants’ (1995: 352). Indeed, the ratings assigned to all quality criteria by TV experts were higher than those assigned by the average combined ratings of conference participants (interpreters, medical doctors, engineers, Council of Europe delegates) (see table 3). Furthermore, even though TV representatives – just like conference participants – gave priority to *sense consistency with original message*, followed by *logical cohesion*, they ranked *pleasant voice* as third most important criterion and *fluency of delivery* as the fourth, followed by *native accent*.

Criterion (rated on a 4-point scale)	TV professionals (n=19)	Conference participants (N=124) [“average” in Kurz (1993)]
1. native accent	2.84 (6)	2.37 (8)
2. pleasant voice	3.47 (3)	2.6 (6)
3. fluency of delivery	3.32 (4)	3.1 (5)
4. logical cohesion	3.68 (2)	3.46 (2)
5. sense consistency with original message	3.84 (1)	3.69 (1)
6. completeness of interpretation	3.53 (8)	3.2 (4)
7. correct grammatical usage	2.79 (7)	2.6 (6)
8. use of correct terminology	3.32 (4)	3.4 (3)
average	3.22	3.06

Table 3. Comparative ratings of quality criteria by Kurz & Pöchhacker (1995: 352).

### 1.4. PÖCHHACKER AND ZWISCHENBERGER (2010)

Within a web-based questionnaire survey on quality and role, Pöchhacker and Zwischenberger (2010) asked the respondents, among other things, to rate nine of Bühler’s “linguistic (semantic)” criteria, plus *lively intonation* and *synchronicity*, on a four-point scale. The questionnaire on simultaneous interpreting quality expectations was administered through *Limesurvey* to professional interpreters (AIIC members), and completed by 704 respondents. As in Bühler’s study (1986), and in Kurz’s (1989; 1993), the criterion *sense consistency with the original* received

the highest ratings, followed by *logical cohesion*. Unlike similar mentioned studies, in this case, the third most important criterion was *fluency of delivery*. The criteria *lively intonation* and *pleasant voice* received almost the same ratings, 28.2 and 27.5 respectively.

Criterion	very important	important	less important	unimportant	N=
Fluency of delivery	<b>70.7 (3)</b> 49	<b>28.6</b> 49	<b>0.7</b> 2	- -	<b>704</b> 47
Correct terminology	<b>61 (4)</b> 49	<b>38</b> 51	<b>0.9</b> -	<b>0.1</b> -	<b>703</b> 47
Correct grammar	<b>54.4 (5)</b> 48	<b>40.4</b> 50	<b>5.1</b> 2	<b>0.1</b> -	<b>701</b> 46
Sense consistency with original	<b>88.3 (1)</b> 96	<b>11.1</b> 4	<b>0.6</b> -	- -	<b>702</b> 47
Lively intonation	<b>28.2 (8)</b>	<b>59.3</b>	<b>11.7</b>	<b>0.9</b>	<b>703</b>
Native accent	<b>14.1 (11)</b> 23	<b>42.1</b> 47	<b>39.7</b> 28	<b>4.1</b> 2	<b>701</b> 47
Logical cohesion	<b>74.8 (2)</b> 83	<b>24.8</b> 15	<b>0.4</b> 2	- -	<b>698</b> 47
Pleasant voice	<b>27.5 (9)</b> 28	<b>58.5</b> 61	<b>12.7</b> 9	<b>1.3</b> 2	<b>702</b> 46
Synchronicity	<b>15.3 (10)</b>	<b>52</b>	<b>30.1</b>	<b>2.7</b>	<b>675</b>
Appropriate style	<b>36.2 (7)</b> 17	<b>55.6</b> 68	<b>7.4</b> 15	<b>0.9</b> -	<b>702</b> 47
Completeness	<b>47.7 (6)</b> 47	<b>45.7</b> 49	<b>6.3</b> 4	<b>0.3</b> -	<b>698</b> 47

Table 4. Relative importance of output-related quality criteria (in percentages). The criteria are presented in the same order as in the web-based questionnaire. Data from the present survey are shown in boldface, while Bühler's (1986) results appear underneath in normal font (Pöchhacker and Zwischenberger 2010). Ranking numbers (in brackets) in the first column are mine.

### 1.5. MEAK (1990)

Meak (1990) devised a questionnaire to be administered to ten specialised medical doctors with a significant experience of participation in international conferences (1990: 8). The questionnaire was designed to find out what a specific audience (medical doctors) expected from a simultaneous interpretation, and to use this information as feedback to “target” the interpretation to “the specific qualities of that audience” (1990: 8). Questions (see appendix) were related to: (1) the ‘effectiveness’ of simultaneous interpretation in medical conferences; (2) the irritating aspects of the interpretation (incorrect terminology, incompleteness or omission of numbers and data); (3) the kind of information on the speaker

required by the audience; (4) the kind of information that is considered necessary and therefore that the interpreter should select; (5) the irritating aspect of speed of speech and its effect on comprehension; (6) the importance of the end of a presentation; (7) the importance of translating acronyms. The author maintained that the results of her survey did not have any statistical value (1990: 13). Responses and comments varied, but in some cases respondents gave similar answers: data and figures were expected to be accurate; descriptions of films needed to be translated; data shown on tables could be selected. A good translation of the end of a presentation as well as of final debates in conferences was also required.

#### 1.6. CHIARO AND NOCELLA (2004)

Chiaro and Nocella (2004) used a questionnaire on quality expectations with the same “linguistic” criteria as those used by Bühler (1986) and partially the same “extra-linguistic” criteria from Bühler for the first web-based survey of its kind. The extra-linguistic criteria they used were: *preparation of conference documents, endurance, ability to work in a team* (also present in Bühler’s questionnaire), *concentration, physical well-being, mnemonic skills, encyclopaedic knowledge, and absence of stress* (not present in Bühler’s questionnaire). The authors pointed out that in their study “the criteria under scrutiny were not measured on an itemised-category scale but on a rank order scale”, because “Bühler’s results showed that interpreters found it difficult to point to unimportant factors” (2004: 283):

The difference between these two types of single-item scales is that responses, options or categories, while on a rank order scale they are required to order a set of objects with regard to a common criterion (Chiaro and Nocella 2004: 283).

In addition, they decided to separate the nine “linguistic” and the eight “extra-linguistic” criteria “on two different ranking-order scales” in order to reduce respondents’ “mental effort” (2004: 283).

The questionnaire was sent to about 1,000 “interpreters belonging to several professional associations” in the world; 286 were returned, mainly from respondents born in Europe (49%) and America (56%); the majority of subjects were female (71%) with an average age of 45 (2004: 286).

Results showed that linguistic criteria were ranked in the following way: *consistency with the original* (1); *completeness of information* (2); *logical cohesion* (3); *fluency of delivery* (4); *correct terminology* (5); *correct grammatical usage* (6); *appropriate style* (7); *pleasant voice* (8); *native accent* (9). The “distribution of the degree of importance given to each linguistic criteria” is shown in the figure 1, below (2004: 287):

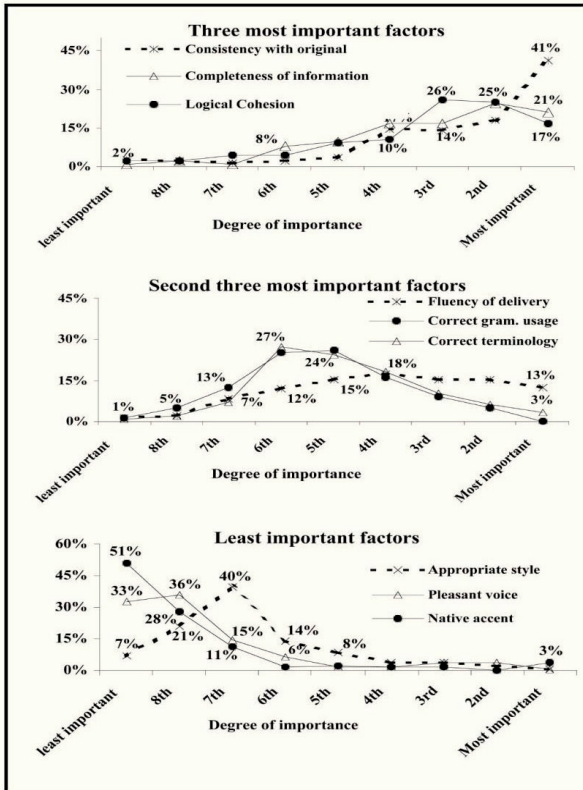


Figure 1. “Distribution of the degree of importance given to each linguistic criteria” (Chiaro and Nocella 2004: 287).

As far as extra-linguistic criteria are concerned, the sum of the scores given to each criterion showed that *concentration* and *preparation of conference documents* were considered the most important; followed by (in this order): *ability to work in a team*, *endurance*, *physical well-being*, *mnemonic skills*, *encyclopaedic knowledge*, *absence of stress* (2004: 288-289). The authors also drew a “perceptual map (i.e., geometric configurations)” of the data on linguistic criteria. Before doing that, the data had to be “explored using Multidimensional scaling”. Therefore, a “scree test”<sup>1</sup> was “performed”, and the result was that “two dimensions were the best solution” for the data; “moreover, the Shepard diagram” confirmed “the model” (2004: 289). Thus, a perceptual map was drawn, called “interpreter’s image of linguistic criteria” (see figure 2), where the horizontal axis was the “discriminating quality axis” and the vertical one was the “structural axis” (2004: 290).

1 “The scree plot is a test performed to decide how many dimensions are used in drawing the perceptual map” (Chiaro and Nocella 2004: 292).



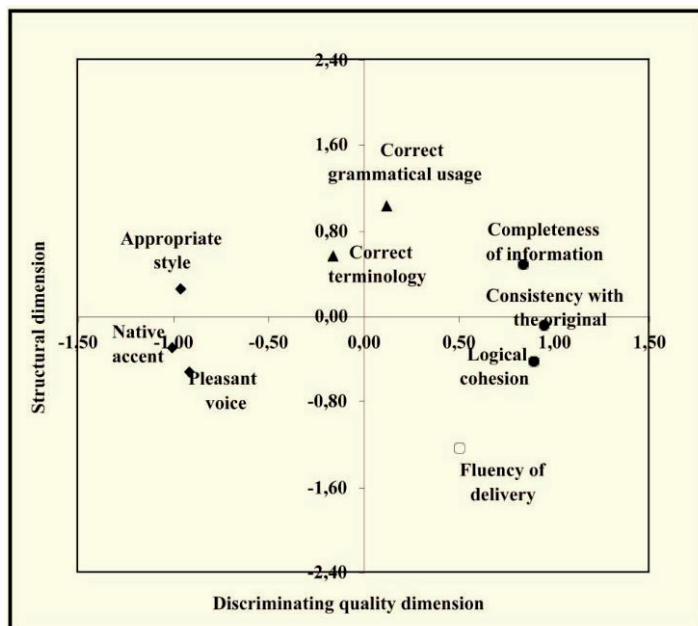


Figure 2: “Interpreter’s image of linguistic criteria”  
(Chiaro and Nocella 2004: 290)

According to the authors, “this chart confirms and strengthens” their “previous findings”; therefore, the criteria *completeness of information*, *consistency with the original* and *cohesion* “score very closely on the right of the discriminating quality axis”, meaning that “they appear to be perceived in a more or less similar way”. The same is true for the criteria *voice quality*, *accent* and *style* on the left (negative) side of the horizontal axis; while the criteria *lexis* and *grammar* “score very closely and positively on the structural axis”. The criterion *fluency of delivery* “appears to stand alone and thus results in being dissimilar to any other features”. The authors inferred that such a position was due the fact that “intonation is considered by interpreters [...] on the interface of the two dimensions” (2004: 290):

On balance, considering that fluency in language plays the double role of both embellishment (i.e., in terms of speed, voice control and absence of hesitation) and structure (i.e., the supra segmental significance of stress, pitch and tone) then it would appear that this feature has indeed been placed where it would most obviously occur (2004: 290).

#### 1.7. KOPCZYŃSKY (1994)

Kopczyński (1994) conducted a survey on quality expectations in different user groups: respondents included those involved in the humanities, experts in science and technology, and diplomats. The questionnaire was administered to “people who attended international conferences as speakers or hosts, or who par-

ticipated in negotiations in one or both of these roles" (1994: 91). Such a variability of subjects was due to the fact that according to the author "quality is not an absolute value, but rather contextually determined" (1994: 88). This means that the variables changed according to the communicative situation. Variables were: the "speaker", his/her "attitude, status and intention"; the "interpreter" and his/her competence; the "receptors" and their "attitude"; the "message", its "form" and "illocutionary force"; the "existing norms of interaction and interpretation of a speech community"; the "setting" (1994: 88). The questionnaire flow started with an open-ended question about what the respondents "considered to be the most important function of interpreting a conference" (1994: 92). The following question was the same as the first one, but this time respondents "were asked to grade the priorities alongside the proposed suggestions": *rendering the general and detailed content of T1; terminological precision; style; grammatical correctness of utterances; fluency of delivery; diction; voice qualities*. The following two questions had the same pattern of the previous ones: the respondent is asked to "mention whatever s/he considered as most irritating" (1994: 92) in an open-ended question first, and then suggestions to rank: *faulty terminology; ungrammatical sentences; stylistic mistakes; incomplete sentences; lack of fluency; poor diction; monotonous intonation and tempo; speeding up and slowing down; too general and too detailed rendition of content*. The remaining five questions concerned "the more or less active role of the interpreter (the ghost role vs. the intruder)" (1994: 92).

The results showed that the criterion *rendition of detailed content* was considered the most important and *terminological precision* the second most important both by "speakers" and "receptors"; while the third most important criterion was *fluency* for "speakers" (followed by *grammaticality* and *style*), and *style* for "receptors" (followed by *fluency*) (1994: 93). As to irritating aspects, both "speakers" and "receptors" agreed in considering *incorrect terminology* as the most irritating feature; the second source of irritation for the "speakers" was *exact rendition of the content*, while for the "receptors" was the item *unfinished sentences*, followed by *grammaticality* (1994: 94). As regards "the role of the interpreter", "all were in favour of empathy with the speaker and considered the ghost role of the interpreter as preferable" (1994: 96); according to the majority of respondents, "the interpreter should imitate the tempo and the intensity of voice of the speaker, but not necessarily the gestures". Although all respondents preferred the ghost role of the interpreter, the majority of them allowed "corrections of the speaker (with some reservation) and additional explanations". "Speakers" accepted "to be corrected", while "receptors opposed the idea" (1994: 96-97).

## 1.8. MOSER (1995)

Peter Moser (1995) reported a "Survey on Expectations of Users of Conference Interpretation" "entrusted" to "SRZ Stadt + Regionalforschung in Vienna/Aus-

tria” by the AIIC Research Committee (1995: 1). The survey was designed taking into account the differences of expectations (i) between interpreters and users, and (ii) among different user groups (as detected by Kurz 1989, 1993). The survey also had to consider the effect of the situational context (conference type) on the evaluation (1995: 4). It was an interview survey with questionnaire; interviews (over 200) were conducted by interpreters at 84 different meetings around the world, while the interviewees were all conference participants (listeners and/or speakers) (1995: 5-7). The questionnaire was made up of 33 questions (1995: 24-47). The first question about quality categories (number 2 in the questionnaire) asked the respondents to rate on a five-point scale what criteria of an interpretation they considered most important among *completeness of rendition, clarity of expression, correct terminology* or *other* (1995: 25-26). The following questions asked the respondent whether – in the conference s/he has just attended – it “is more important that the interpreter focuses on the essentials or gives a complete rendition” and “whether fidelity to the meaning of the original or the literal reproduction of what is said is more important” (1995: 26-27). The fifth question asked respondents to “indicate” (on a five-point scale) how important the following criteria were considered: (“the interpreter should...”) *speak in a lively and animated way; speak in complete sentences; interpret titles, names of functions, positions, offices held; interpret subtitles in graphs and tables on transparencies or slides; interpret abbreviations; anything else* (1995: 27-28). The sixth question was divided into two parts: the first part was an open-ended question about what most “irritates” the respondent; the second part suggested a series of criteria, asking “how irritating is an interpreter who”: *makes long pauses, lags far behind the original; speaks very quietly; speaks in a monotonous way; ums and ahs; has a foreign or regional accent* (1995: 28-30). An interesting aspect of the data processing is the following:

All the open-ended replies were first sorted into a highly differentiated category matrix. Each category was then allocated a code so that subsequently the open-ended replies could be statistically analyzed in conjunction with the replies to the closed questions (1995: 49).

In this way, if a respondent spontaneously mentioned the importance of one criterion two or more times, that number of references appeared in the statistics; therefore, there was no one-to-one correspondence between the number of interviewees and the number of references to a criterion. Quality criteria, both spontaneous and suggested by the questionnaire, were grouped into four general categories: *content* (at least *faithfulness to the original*); *synchronicity* (referring to the need for a short *décalage*); *rhetorical skills* (*regular delivery, absence of hesitation, complete and grammatically correct sentences, clarity of expression*); *voice* (*lively, non-monotonous voice, clear enunciation, ‘a natural sounding voice’ or ‘an agreeable voice’*) (see figure 3) (1995: 8).

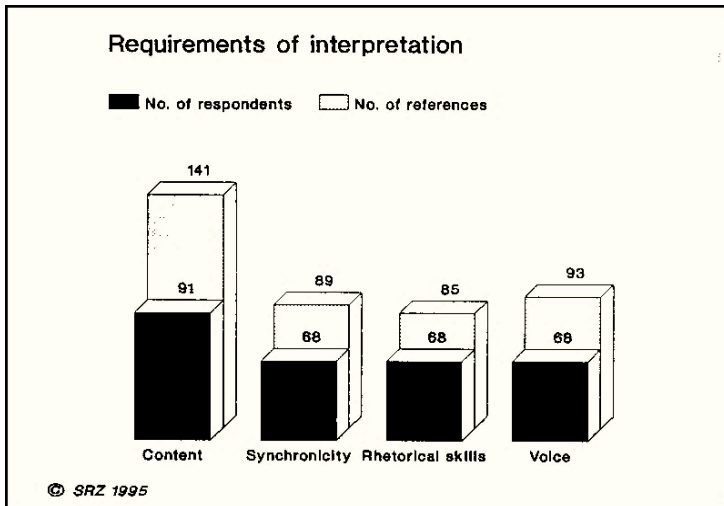


Figure 3. “Requirements of interpretation” (Moser 1995: 8).

Data showed in figure 3 varied according to the conference type (“small/large general/technical meeting”), gender, level of respondents’ experience in conference interpreting (“newcomers, less experienced, very experienced”).

## 2. QUALITY ASSESSMENT QUESTIONNAIRE-BASED SURVEYS IN SIMULTANEOUS INTERPRETING

This section deals with a literature review of studies on quality assessment of simultaneous interpreting (SI), i.e. a questionnaire-based survey on the judgment of real interpretations, on the basis of pre-definite or spontaneous criteria. These quality criteria are linked to the items of the questionnaires, requiring the subjects either to rate each of them on a numeric scale, to rank them in order of importance or to comment on them through open questions. Subjects are interpreters, interpreting students, delegates, other types of users. Some of the studies reviewed in this section, beside questionnaires on quality assessment also include questionnaires on quality expectations, administered either before or after questionnaires on assessment, with the aim of comparing the results of the two evaluation approaches. Only one of the studies reported (Collados Aís *et al.* 2007) also includes a “contextualization” questionnaire, designed to better elicit the respondent’s personal understanding of the criterion the subject was required to assess.

### 2.1. GILE (1990)

Gile (1990) carried out a survey (a “case-study”) on quality assessment (judgement) of SI, with the aim of studying “the subjective perception of quality by delegates” (1990: 68). A bilingual questionnaire was devised to be administered to French and English speaking delegates at a conference. Eighteen French speaking delegates and five English speaking delegates returned the questionnaire (1990: 67). The quality was assessed according to the following criteria, rated on a five-point scale: *general quality*, *linguistic output quality*, *terminological usage*, *fidelity*, *quality of voice and delivery* (voix, rythme et intonation), *main weaknesses of interpretation* and *general comments on interpretation* (1990: 71). The results showed that assessment of general quality was quite homogeneous, since it was considered “good” or “very good” by almost all the respondents. Homogeneity also characterised all the other criteria, with the exception of *voice*, which presented the most varied judgements (see table 5, where the overall results are reported). Moreover, ratings assigned by English speaking delegates were higher than those assigned by French speaking delegates, since the former rated as “very good” all the criteria and defined the interpretation as “excellent”, without any “weakness” (1990: 67-68, 71). Six out of the eighteen French speaking delegates judged *quality of voice and delivery* with the lowest ratings; however, this evaluation had no impact on the judgement of the *general quality of interpretations* of the interpretation, since the rating of this criterion was in line with those of other criteria (1990: 68).

CRITERIA	Very poor	Poor	Medium	Good	Very good
General quality of interpretation			9%	45%	45%
Linguistic output quality			14%	45%	41%
Terminological usage			18%	50%	32%
Fidelity			9%	50%	41%
Quality of voice and delivery		9%	22%	43%	26%

Table 5. “General presentation of results” is in Gile (1990:70)

### 2.2. NG (1992)

Ng (1992) designed a quality assessment questionnaire to be administered to Japanese native speakers, with the following working hypothesis: “appropriate use of speech levels is significant in English-Japanese interpretation at the conference level” (1992: 36), because speech levels could be loosely defined as “grammaticised social relationship indicators” (1992: 35). “Loosely defined, speech

levels refer to the choice of expressions which indicate the speaker's social and psychological stance at a particular point in time" (1992: 36). "Speech levels" were otherwise referred to as "levels of politeness" (1992: 36, 37). The questionnaire was administered to ten native Japanese speakers and submitted in two stages: in the first stage assessors were asked to indicate the general usability of the interpretation; in the second stage, they were asked to indicate if speech levels "interfered" with the content delivery. Specifically, in stage 1 subjects were requested to listen to the students' prerecorded English-Japanese interpretations, executed by native English interpreters (not Japanese), and then answer three questions about: (1) whether or not the interpretation "could be followed", and the message the interpreter was trying to convey could be caught; (2) the general impression of the Japanese used by each interpreter; (3) which of the interpreters gave the best interpretation and which gave the worst, if possible explaining the reasons for the ranking. Stage 2 consisted of an individual interview held approximately a week after Stage 1. After a short briefing on "the technical linguistic terms used to describe speech levels in Japanese", with the aid of "examples", the subject listened to "short segments of the tapes" to "refresh their memory". After listening, they were asked 4 questions on: (1) what they thought of the "inappropriate use of speech levels in interpreting"; (2) whether that "interfered with the content delivery"; (3) whether they found that "offensive"; (4) whether that made them feel uncomfortable. Lastly, the subjects were asked to rate (on a five-point scale) the performance of each interpreter relative to "the control of Japanese speech levels" (1992: 37). The results showed that answers relative to Stage 1 "were general and covered a broad area". However, the comments "fall into three broad categories of content, language and extralinguistic criteria". "Comments on the language" [...] could be referred to the following "subcategories": (2a) "naturalness, e.g. intonation, pronunciation and accent"; (2b) "grammatical structure"; (3c) "choice of vocabulary"; (2d) "speech levels". "All the subjects were particularly concerned with whether the interpreters had grasped the meaning of the message", therefore the interpreters "were criticised for giving obscure interpretations". In the end, the results from Stage 1 showed that "though speech levels were discussed, it was by no means the most important variable singled out by the subjects". The same aspect characterised findings in Stage 2, where the subjects found it difficult to "isolate speech levels as a separate variable"; in some cases they "seemed to stray from the topic of speech levels to the more general problem of fluency". The general result of the study was that for native Japanese speakers "appropriate use of speech levels" was "important", but their "misuse" did not necessarily result in offending the audience" (1992: 37-40).

### 2.3. MARRONE (1993)

Marrone (1993) carried out a survey on “audience expectations and preferences” in consecutive interpreting, with particular attention to “definition and evaluation of interpretation quality” (1993: 35). The questionnaire was administered to an audience of 150 users (mostly students, only a dozen of them were researchers and professors) of consecutive interpreting who attended a lecture on constitutional law. The questions aimed at eliciting information about “audience’s preferences and reactions” (1993: 35). The questionnaire was made up of seven questions designed to elicit the respondents’ reactions to the interpretation of segments, the speed of delivery, the effect of the interpreter’s tiredness on performance, the translation of institutional terms, the interpreter’s role as a translator or a mediator.

The questionnaire was administered after the performance of a consecutive interpretation, but from the wording of the questions it is not clear if this asked the respondent to refer to the performance of the interpretation for an ‘assessment’; on reading the questions it seems that the performance served as an input to indicate “preferences” about the ‘ideal consecutive interpretation performance’.

Results showed that respondents “seem to attach far more importance to substance, fidelity, completeness of information than to the linguistic quality of prosodic features of interpretation”, even if “scores related to such features tended to fluctuate widely” (1993: 38). Moreover, responses “indicate that it is appropriate” that interpreters “attempt a degree of cultural mediation”.

Parameters	Score
Complete transmission of the original message	216
Quality of style and correct terminology	165
Quality of intonation and delivery	137

Table 6. Partial results by Marrone (1993, in Soleer Caamaño 2006: 75).

### 2.4. VUORIKOSKI (1993)

Vuorikoski (1993) conducted a survey on quality perception of users of SI from English into Finnish with the aim of studying the effectiveness of communication mediated by interpreters in conferences held in Finland. A questionnaire was administered to the attendees of 5 seminars; 177 participants responded, some of whom were later interviewed by phone to complete the information provided in the questionnaires. The part of the questionnaire regarding the quality assessment included absolute (yes/no) questions relative to the following aspects: the interpreter was well informed; the interpretation was coherent and easy to follow; it was accurate; rhythm was pleasant; it was fluent; terminology



was correct. The following question asked respondents to rank the mentioned criteria. The results show that all the questions received a “yes” answer in a percentage equal to or higher than 50%. As to the ranking, the good preparation of interpreters was ranked first, coherent and easy-to-follow interpretation was ranked second and fluency of interpretation was ranked third (Vuorikoski 1993, in Soleer Caamaño 2006: 77-79).

Criteria	Assessment	Ranking
Interpreter well informed	67%	22.7% (1)
Coherent and easy to follow interpretation	85%	19.7% (2)
Accurate interpretation	50%	16.9% (4)
Pleasant speech rhythm	64%	8.6% (6)
Fluency	50%	18.6% (3)
Terminology	60%	13.5% (5)

Table 7. Results from quality assessment and ranking of criteria in Vuorikoski (1993, in Soler Caamaño 2006: 79).

## 2.5. MACK AND CATTARUZZA (1995)

Mack and Cattaruzza (1995) conducted a survey on user reception and expectation of SI based on the methodology adopted by Vuorikoski (1993) for a similar survey in Finland. According to this “multimodal” research, “data obtained through the questionnaire were integrated and partially checked by telephone and personal interviews as well as by non-reactive research in the conference situation” (1995: 38). The aim of this survey was to study “how quality is measured” in order to explore the possibility “to go beyond purely subjective judgment” (1995: 37). The questionnaire was distributed at 14 conferences, but only five of them were selected because their context was considered to match the criterion of “users’ high communication needs” (1995: 39). The number of participants who returned the questionnaire completed was 75; of these, only the 58 Italian participants were considered for analysis (1995: 41). Two central questions asked the respondents to evaluate – through a rating on a five-point scale – the simultaneous interpretation heard (assessment), and then to “indicate the importance of the criteria listed (their wishes and expectations)” (1995: 43). The criteria were the same as those used by Vuorikoski (1993) in her survey: the interpreter was well informed; interpretation was coherent and easy to follow; it was accurate; rhythm was pleasant; it was fluent; terminology was correct.

Findings in assessment (factual experience) showed that participants found that the interpretations were easy to follow, had a pleasant speech rhythm and that the interpreters were well informed – the mean rating of all these criteria



was 4.2; the mean rating for fluency was 4.1, while both terminological correctness and accuracy received a rating of 3.8 (“slightly lower”). As to the results of quality expectations, terminological correctness received the highest mean rating (4.5), followed by the preparation of the interpreter (4.3) and an accurate and easy-to-follow interpretation (both 4.1) (1995: 43-45).

All in all, the average values of the ratings moved in a rather narrow range (quality experienced: 3.8 to 4.2; expectations: 3.8 to 4.5). The largest deviation between experience and expectations appeared within the criterion of terminological correctness (-0.7), followed by accuracy (0.3) and informedness (-0.1), while the ratings for quality experienced were higher than those for expectations in the characteristics ‘easy to follow’ (+0.1), and ‘fluent and pleasant speech rhythm’ (+0.3) (Mack and Cattaruzza 1995: 45).

Criteria	Assessment / “experience”	Expectations
informed	4.2	4.3
easy to follow	4.2	4.1
accurate	3.8	4.1
pleasant rhythm	4.2	3.9
fluent	4.1	3.8
terminology	3.8	4.5
average	4.05	4.1

Table 8. Comparison of ratings of quality assessment and expectations in Mack and Cattaruzza (1995, in Soler Caamaño 2006: 89).

## 2.6. DOERFLINGER (1993)

Doerflinger (1993), on behalf of the Directorate General for Interpretation (SCIC) of the European Union, reported a survey on quality assessment of the conference interpreting service provided by the SCIC: “*qualité de l’interprétation et qualité globale du service fourni*” (2003: 173, italics in original). The notion of *global quality* here was referred to the interpretation as the *result* of a process starting with recruitment, and ending with planning, providing of documents etc. (2003: 173). The questionnaire was administered to interpreters and delegates in 80 conferences of different types. 800 delegates and 700 interpreters returned the questionnaire (2003: 175). The questions were relative to the following criteria (in brackets the percentage of respondents who answered ‘yes’, or who agreed with the questions): *l’interprétation est satisfaisante* (91%); *le message passait bien* (80%); *la terminologie utilisée par les interprètes était appropriée* (78%); *l’expression des interprètes était agréable (voix – ton – débit – volume)* (78%); *les sujets étaient bien maîtrisés* (74%); *le professionnalisme des interprètes paraissait satisfaisant* (81%) (2003: 175).

## 2.7. GARZONE (2003)

Garzone (2003) carried out a survey-based research – a “small-scale pilot study” (2003: 25) – on both quality expectations and quality assessment. Sixteen subjects were administered a short questionnaire consisting of

four of the criteria of quality used by Kurz in her surveys (Kurz 1988; Kurz 1993 based on Bühler 1986), two concerning content and two concerning form: *fluency of delivery*; *pleasantness of voice*; *logical coherence of utterance*; *sense consistency with the original message*, which was reformulated as *fidelity to source text* in order to make it more readily comprehensible for the layman (Garzone 2003: 26).

The subjects were “eight doctors and other professionals (mostly engineers), operating in different technical fields” (2003: 26). The respondents first had to rate the criteria according to their quality expectations; then, listen to two interpreted versions of the same text to assess them according to the same criteria. One version, the “original” interpretation, was “correct and characterised by pleasantness of voice and good fluency” (2003: 27). The other version was artificially manipulated (“fabricated”), because it was re-recorded by another interpreter who “improved the rendition of the source text [...], but adding a number of hesitations” (2003:26) and altered the prosody to make it “somewhat erratic” (2003: 27); therefore this version was “at least as correct [as the first one] but more objectionable formally (2003: 27). For both expectations and assessment questionnaires “the rating scale was from 1 to 10, the easiest for Italians as it used in Italian schools for assessing students’ performances” (2003: 26). Questionnaire delivery was followed by “short interviews” with the respondents (2003: 27).

The results showed that in the expectation questionnaire the criteria *pleasantness of voice* and fluency of delivery were considered “less important” than other criteria; while in the assessment questionnaire it was clear that they had a “marked impact on their assessment of other aspects of the two performances” (2003: 27) (see table 9). In the light of these results, the author put forward two considerations. The first was that the interviews made after the completion of the questionnaires confirmed that the ratings assigned to voice quality and fluency were “ideologically biased”, because they were “based on the idea that after all form is not important, what really counts is content” (2003: 28). The second consideration was a consequence of the first one:

When one speaks in purely abstracts terms the evaluation of each single criterion is given in isolation, while in real evaluation processes the different elements overlap and interfere with one another: the performance that is poor in terms of prosody and fluency is perceived as less correct and less coherent, even when in actual fact it isn't. This may be due either to objective difficulty in following the sense of an oral text which is not well presented, because it seems less comprehensible, or to a “psychological effect” which makes the interpretation in Tape B appear less reliable (Garzone 2003: 28).

EXPECTATIONS			
Criteria	Doctors	Other professionals	Average
Pleasant voice	6.12	6.25	6.18
Fluency of delivery	6.12	5.62	5.87
Fidelity of ST	9.00	8.87	8.93
Coherence of utterance	8.87	8.5	8.68
ASSESSMENT			
Criteria	Doctors	Other professionals	Average
<b>Tape A</b>			
Pleasant voice	8.50	7.87	8.18
Fluency of delivery	9.12	8.12	8.61
Fidelity of ST	8.50	8.75	8.62
Coherence of utterance	9.12	8.87	8.99
<b>Tape B</b>			
Pleasant voice	6.12	5.87	5.99
Fluency of delivery	4.5	4.25	4.37
Fidelity of ST	5.62	5.12	5.36
Coherence of utterance	6.00	5.50	5.57

Table 9. Results of the survey on SI quality expectations and quality assessment (Garzone 2003: 26-27).

## 2.8. Russo (2005)

Russo (2005) for the first time studied both assessment and expectations of quality by different user groups in the simultaneous interpretation of films. The paper is based on the findings of the studies conducted by Guardini (1995) and Palazzini Finetti (2000). Both studies include questionnaire-based surveys based on Guardini's questionnaire, being the questionnaire by Palazzini Finetti "a modified and slightly shortened version of Guardini's, to allow comparison of results" (Russo, 2005: 7). Guardini's survey examined the simultaneous interpretation of films from English into Italian performed by professional interpreters; while Palazzini Finetti's survey examined the simultaneous interpretation of films from Spanish into Italian executed by interpreting students (Russo, 2005: 7). The questionnaire was administered to interpreting students, film critics and other attendees (clerks, lawyers, doctors, etc.) of two film festivals held in Italy who used the film interpreting service; in total 195 audience members returned

the questionnaire. Questions relative to quality assessment constitute the first part of the questionnaire, while the second part is dedicated to the quality expectations (preferences). The quality assessment criteria were: *general quality, voice quality, formal and grammatical correctness, delivery quality, fluency, synchronisation (image-dialogue), dialogue completeness and expressiveness*. The question on quality expectations asked the respondents which of the following features of SI of film they considered important: *dialogue completeness with all details, rendition of the general dialogue content, synchronisation, acting, fluency, pleasant voice, adequate style, explanation of non-verbal elements, for ex. road and shop signs, written messages..., other* (Guardini 1995: 23-26; in Russo 2005).

In the assessment questionnaire relative to the survey conducted by Guardini (1995; in Russo 2005), which included the criterion *general quality*, the criteria that received the highest ratings by film critics was *general quality* (2.9), followed by *voice quality, grammatical correctness* and *word/image synchronisation* (all 2.8), and *fluency of delivery* (2.7). The students assigned the highest rating to *voice quality* (2.8), followed by *general quality* and *grammatical correctness* (both 2.4). The overall average of ratings show that *voice quality* was the criterion mostly appreciated, followed by *overall quality* (2.8) and *style* and *fluency of delivery* (both 2.5) (Russo 2005: 12) (see table 10). However, in both Guardini's (1995) and Palazzini Finetti's expectation questionnaires (in Russo 2005), completed after the assessment ones, the user groups' average rating shows that in both questionnaires the criteria which received the highest ratings were *rendition of general dialogue content*, followed by *fluency of delivery* (Russo 2005: 15) (see table 11).

Aspect of quality	Overall average (n=84-90)	Critics (n=8-10)	Students (n=17-18)	Other (n=48-53)
General quality	2.8	2.9	2.4	2.9
Voice quality	2.9	2.8	2.8	2.9
Style	2.5	2.5	2.1	2.7
Grammatical correctness	2.7	2.8	2.4	2.6
Fluency of delivery	2.5	2.7	2.1	2.6
Word/image synchronisation	2.4	2.8	2.1	2.3
Dialogue completeness	2.4	2.4	2.2	2.4
Expressiveness	2.2	2.2	2.1	2.3

Table 10. "Ranked average ratings (4-point scale) of quality [assessment] features by user group (based on Guardini 1995)" (Russo 2005: 12).

Feature of quality	Guardini (1995)	Palazzini Finetti (2000)	Kurz (1993)
Rendition of general dialogue content	3.4 (1)	3.6 (1)	3.7
Fluency of delivery	3.2 (2)	3.2 (2)	3.1
Word/image synchronisation	3.1 (3)	3.0 (4)	–
Dialogue completeness with all details	2.8 (4)	2.5 (6)	3.2
Adequate style	2.7 (5)	3.1 (3)	–
Pleasant voice	2.7 (5)	2.8 (5)	2.6
Acting	2.6 (7)	2.5 (6)	–
Explanation of nonverbal elements	2.3 (8)	2.2 (8)	–

Table 11. “Comparison of quality-related preferences in film and conference interpreting (ranked average ratings of importance on a 4-point scale)” (Russo 2005: 15).

## 2.9. CATANA (2005)

Catana (2005) in her MA dissertation (unpublished) studied the impact of some specific features of voice in SI quality assessment. The objective of her study was to analyse how Italian intonation and diction influenced users’ judgments of SI quality. The questions in the survey were the following (rated on a five-point scale): *overall quality* assessment; *professionalism* of the interpreter; *credibility* of the interpreter; *sympathy/captivation* caused by the interpreter; *pleasant listening*; *voice* (according to volume, speech rate, intonation, tone); *overall assessment of voice* (from “very harmonious” to “not harmonious”); aspects of *diction* (if any) that might had influenced assessment: open-ended question followed by rating of diction followed by a multiple choice (native accent, expressivity, intonation, stressed vowels; rhythm); self-evaluation of *comprehension* of text; ranking of the features considered for the overall evaluation of quality (native accent, pleasant voice and intonation, fluency, logical cohesion, correct grammar, transmission of general sense; correct terminology; style; correct diction and articulation); the last question (open-ended) was addressed to subjects who had obtained a MA degree in interpretation and is about the qualifications of a conference interpreter. The questionnaire was administered to 30 MA interpreting students. The subjects were divided into 3 groups of 10 respondents, in order to evaluate 3 different versions of an interpretation (from Spanish into Italian): the first version with monotonous intonation (intonation was made monotonous by the speaker, who was not a real interpreter and read an interpretation) and correct execution of acute/grave accents of “e” and “o”; the second version with neutral intonation and correct execution of accents of “e” and “o” (read by a professional interpreter); the third version with neutral intonation but incorrect (reverse) execution of acute/grave accents of “e” and “o” (read by the same interpreter of the second version); a completely neutral control version, listened to by all the groups.

According to the results of the survey (2005: 117-202), the monotonous intonation negatively influenced the rating of the *overall quality* of interpretation, the *professionalism*, the *credibility* of the interpreter and his *ability to captivate the audience*. The overall judgment of *voice* and the *pleasantness of listening* were also affected by the monotonous intonation, as was the self-evaluation of *comprehension*. As to the indication (self-evaluation) of the aspects that influenced the assessment of the *overall quality*, for the first version (monotonous intonation) the features that most influenced the evaluation were fluency and correct terminology (ranked first, on average), followed by style; for the second version the ranking was the following: transmission of general sense, followed by fluency and correct terminology; for the third version the ranking was: transmission of the general sense, followed by pleasant voice and intonation and correct diction and articulation. The global evaluation showed that volume and speech were “medium”, intonation was “monotonous” and tone “neutral”. Nonetheless, the analysis of the individual ratings assigned to each sub-parameter relative to the voice showed that the subjects recognised that the first version was more monotonous than the second and third; however, third version was considered “melodious” compared to the second one. As to the *diction*, with respect to the first version 80% of respondents found no marked feature in the interpreter’s diction; in the second version, 90% of subjects maintained that there were marked features, since “intonation”, “speech rate” and “expressiveness” were considered “clearly perceptible”; in the third version, 100% of respondents found marked features, mainly the open vowels “e” and “o” (as was actually the case) (2005: 117-202).

#### 2.10. COLLADOS AÍS ET AL. (2007)

Collados Aís et al. (2007) published a research study carried out by the ECIS group (Quality assessment in simultaneous interpreting), coordinated by Collados Aís. This research followed the study by Collados Aís (1998) on the influence of intonation in the quality assessment of SI. The aims of the research were: (i) to analyse quality expectations of users on the basis of “eleven parameters”; (ii) to analyse “conceptualization and its incidence on the quality evaluation of a SI executed by specialised users” (2007: 6). The 11 parameters tested in both the expectation and the assessment questionnaires were: *accent*, *pleasant voice*, *fluency*, *logical cohesion*, *correct transmission of the original message*, *complete transmission of the original message*, *style*, *intonation*, *diction*, *correct grammar*.

For the assessment session, each quality parameter was artificially manipulated, intervening in the text or in the performance of the interpretation. In total, 12 videos were created for the assessment questionnaire: one for each parameter plus one control video, where the interpretation was not manipulated. As to the manipulations: *intonation* was made “more monotonous”, *pleasant voice* “more nasal”, *style* “more guttural, then less concise”; in *diction* the “articulatory

ability was reduced”; in *fluency* “speech rate was modified” and “false starts and self-repairs” were introduced; *accent* was made “native German”, and in *correct grammar* some “most common mistakes” made by “native German speakers speaking Spanish” were introduced; in *terminology* elements of “judicial phraseology” were “replaced with common language words”; in *logical cohesion* “logic discourse markers were removed” or logically “inverted”; in *correct and complete transmission of the original*, “information units were distorted or removed respectively” (2007: 12-13).

The questionnaire for the assessment, in addition to the 11 questions relative to the parameters, also included questions on the *overall quality* of the interpretation, the evaluation of the *original text*, impressions of the *reliability* and the *professionalism of the interpreter* (2007: 14). The subjects were 197 university teachers of law for the expectation questionnaire and 164 equivalent (but not necessarily the same) teachers. The questionnaire on “conceptualization” of parameters was made of 3 parts: the first part included two open-ended questions on one parameter; the second part had to be answered after the subject watched the video relative to the manipulated parameter, and asked the respondent to redefine the parameter after the screening and then rate it on a five-point scale; the third part contained a further evaluation, similar to that in the second part (2007: 16, 255). The subjects of this questionnaire were 32 university teachers of law (2007: 15).

The findings confirmed the hypothesis of the survey: expectations did not match experience (assessment). Manipulation of “verbal” parameters were not easily detected, while the ones relative to “non-verbal” parameters were all detected. Among the “verbal” parameters, the video where *logical cohesion* had been manipulated received the lowest score. The video where the parameter *fluency* had been manipulated received the lowest score (3) in the parameter *correct transmission of the original*; however, this score was lower than that the same parameter received in the video where the *correct transmission of the original* had been manipulated (4.33). *Fluency* also had an impact on *diction* and on *logical cohesion*; in both cases it was also the other way round: ...*logical cohesion* influenced the perception of *fluency* and *diction*. *Fluency* also proved to be interrelated... with *correct and complete transmission of the original message*. Other “combinations of interrelations” were: *style* and *intonation*; *style* and *pleasant voice*; *diction* and *intonation*; *diction* and *pleasant voice*; *complete transmission* and *cohesion*; *terminology* and *correct grammar*. The video where the parameter *style* had been manipulated received a higher score than the control video (2007: 220-221). *Intonation*, as well as all other “non-verbal” parameters, influenced the perception of the overall quality of the interpretation (2007: 173).



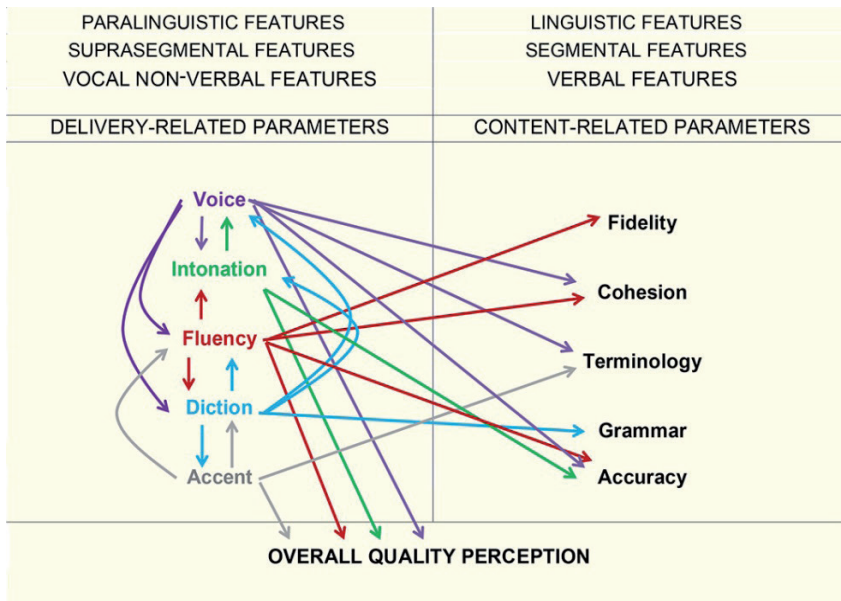


Figure 5. Interdependence of “non-verbal parameters” and their “interrelation” with “verbal” parameters, as found in Collados Aís *et al.* (2007) – the figure is mine.

### 2.11. GARCÍA BECERRA (2013)

García Becerra (2013) in her doctoral thesis studied the impact of the first impression on quality assessment of SI. Both expectation and assessment questionnaires included questions about “formal aspects” (*diction, intonation, fluency, voice, etc.*); “content aspects” (*cohesion, style, terminology, etc.*); “fidelity aspects”; the ranking of those aspects; their variation as a function of the age, sex, etc. of the interpreter; other aspects that influence quality evaluation of expectations. The assessment questionnaire also included questions about a possible impression of the respondent about the interpretation and the interpreter (*competence, confidence, expressiveness, anxiety, pleasantness*) that might or might not have influenced the overall assessment of the interpretation (2013: 651, 654).

The subjects of the survey were university teachers, users of Facebook and Translation and Interpreting students. As to the expectations, university teachers and users of Facebook rated the parameters in a similar way as in Bühler (1986) and in Kurz (1989, 1993), while Translation and Interpreting students assigned higher rates to “formal aspects” (even higher than *fidelity*) (2013: 564). As to findings relative to quality assessment, ratings differed from those assigned to the same parameters in the expectation questionnaire (2013: 564-565). Subjects confessed that their expectations could have been influenced by the “interpreter’s age, gender and vocal aspects”. In fact, the female interpreter obtained more posi-



tive opinions than the male interpreter from all participants, as well as the highest rates in assessment (2013: 566). University teachers and users of Facebook assigned higher rates to the “formal aspects” in the assessment questionnaire than they did in the expectation questionnaire (213: 567). A “high proportion of subjects” admitted “that their first impression” influenced the “overall assessment of the interpretations”; this was confirmed by the “correlational analysis”, which suggested that the “formal aspects” were at the basis of the “impression” (2013: 567). This analysis also suggested that the administration of the survey might have influenced the subjects’ impressions, mainly the sequence of the three interpretations listened to, their “comparison” and the subjects’ “tiredness” (2013: 569). The analysis of the “assignment of adjectives” confirmed the hypothesis on the “perception of the interpreters’ vocal features”. All three groups of subjects defined the male interpreter as “skilled”, “inexpressive” and “insecure”, while the female interpreter was defined as “skilled”, “expressive” and “self-confident”. The “perception of inexpressiveness” might be due to the “frequency range”, the “skillfulness” to the “speech rate” and “silence rate” (2013: 570). “Skillfulness” and “self-confidence” might have “positively influenced” the assessment of *fidelity*, *intelligibility*, *content* and even the *overall quality* of the interpretation, while the perception of “insecurity” might have negatively influenced the assessment of *content*, *fidelity* and *overall quality* (2013: 571). The author concluded that “it looks as if insufficient formal aspects could eclipse remaining parameters in the evaluation mechanism of subjects” (2013: 571).

### 3. GENERAL CONSIDERATIONS ON THE RESULTS OF PREVIOUS STUDIES ON SI QUALITY EXPECTATIONS AND QUALITY ASSESSMENT.

Studies on both quality expectations and quality assessment of SI mainly present the same “linguistic” criteria devised by Bühler (1986), sometimes with similar names, sometimes with the same criteria grouped into other categories, other times with other criteria adapted to the objective of the study (Soler Caamaño 2006: 101; García Becerra 2012: 55, 74, 84) (see Appendix). In general “linguistic” criteria have been divided into ‘content-’ and ‘delivery-related’ or ‘linguistic’ and ‘paralinguistic’ aspects. The results of expectation surveys show that subjects tend to assign higher ratings to content-related criteria like *sense consistency with the original message*, *logical cohesion*, *correct terminology* and lower ratings to form-related criteria, like *style*, *pleasant voice*, *accent* (Bühler 1986; Kurz 1989, 1993; Kopczyński 1994; Chiaro and Nocella 2004; Moser 1995; Pöchhacker and Zwischenberger 2010). In some of these cases (Kopczyński 1994; Chiaro and Nocella 2004; Pöchhacker and Zwischenberger 2010) *fluency* was considered the most important among the form-related criteria; even though it ranked third, after content-related criteria. The only case in which the criterion *pleasant voice* ranked third, after *sense consistency with original message* and *logical cohesion*, was

that in the quality expectations in television interpreting (Kurz and Pöchhacker 1995: 352), and, specifically, in the case where the subjects were TV professionals, because conference participants assigned *pleasant voice* a low rating, in which it ranked sixth. It is worth to mention the ‘uniqueness’ of the AIIC survey reported by Moser (1995), where quality criteria were grouped in four categories (*content, synchronicity, rhetorical skills* and *voice*) and the statistical analysis was made considering all the references (mentions) to the single criteria made by respondents (including the spontaneous ones) along the questionnaire flow. In this case, the results show that, among the “requirements of interpretation”, “references” under the list of the category *content* ranked first (141), followed by references to *voice* (93), *synchronicity* (89) and *rhetorical skills* (89), *rhetorical skills* being the category that grouped *regular delivery, absence of hesitation, complete and grammatically correct sentences, clarity of expression* (Moser 1995: 8).

The general trend observed in the results of the SI quality expectation surveys also applies to the survey on SI quality assessment (Gile 1990; Marrone 1993; Vuorikoski 1993; Mack and Cattaruzza 1995; Garzone 2003; Russo 2005; Catana 2005; Collados Aís *et al.* 2007; García Becerra 2013). Nonetheless, among these studies, those including a quality assessment questionnaire plus a quality expectation questionnaire show that ratings assigned to form-related criteria in assessment are higher than the ratings assigned to the same criteria in expectations (Mack and Cattaruzza 1995; Garzone 2003; Russo 2005; Collados Aís *et al.* 2007; García Becerra 2013). This is even more true in the case of film interpreting (Russo, 2005), where the average rating assigned to voice quality (2.9) is even higher than that assigned to general quality (2.8) (Guardini 1995 and Palazzini Finetti 2000 in Russo 2005). The results of the study by Ng (1992) show that the subjects found it difficult “to isolate speech levels [levels of politeness, in Japanese] as a separate variable”, in some cases the respondents “seemed to stray from the topic of speech levels to the more general problem of fluency” (Ng, 1992: 39), and that “all the subjects were particularly concerned with whether the interpreters had grasped the meaning of the message” (Ng, 1992: 37) even if they interpreted from their mother tongue (English) into Japanese (Ng, 1992: 40). In the study by Mack and Cattaruzza (1995) one of the quality criteria used was *easy-to-follow interpretation*, a parameter that goes beyond the boundaries between content- and form-related aspects, what may be termed a ‘supra-parameter’: in assessment this criterion was assigned a rating that (4.2) which was the same assigned to pleasant rhythm (4.2) and quite close to that assigned to fluent interpretation (4.1) (Mack and Cattaruzza, 1995 in Soler Caamaño, 2006: 89). The fact that the manipulation of aspects like *voice, intonation* and *fluency* had an impact on the assessment (rating) of *cohesion, accuracy, terminology, and overall quality* in Collados Aís *et al.* (2007), might be due to the fact that in actual perception the distinction between content- and form-related parameters is not clear or well defined. This is also evident in the study by Garzone (2003), where the manipulation of *fluency* and *intonation* had

a negative impact on perception, and therefore on the rating of *fidelity of source text* and *coherence of utterance*.

Soler Caamaño (2006), in her doctoral thesis on quality in specialised interpreting training, studied the evaluation criteria of an examination board of a postgraduate course in medical interpreting – a case study. The author analysed the board’s spontaneous (oral) deliberations (transcribed) and identified 67 different quality indicators. After dividing evaluations between “positive” and “negative”, it was found that 13 indicators were used only in negative evaluation:

la falta de reacción al factor sorpresa, la imprecisión en los datos, la repetición de los errores del orador, la mala actitud física en la cabina, la falta de idiomática, la falta de rapidez o agilidad mental, no captar nombres propios (fármacos, siglas), mostrar fallos de cultura general, cometer errores en los datos, respirar de forma que transmita estrés, cometer *Lapsus linguae*, no finalizar las frases, o no hacer un uso correcto del micrófono (2006: 278-279).

By contrast, only 2 indicators were used in positive ones: “*hacerse suyo el discurso original*” and “*ser fiel al contenido*” (2006: 278); the indicator “*voz*” was the only one more frequently mentioned in its positive sense (2006: 279). All the indicators were grouped in six categories:

Por orden de importancia y según frecuencias: prestación, control de la situación, conocimientos específicos, conocimientos generales, competencia traductora, y tácticas y estrategias. Estas categorías no sólo mantienen una relación de vasos comunicantes, sino también una relación jerárquica y de dependencia, de parte y condición de la categoría superior (2006: 280).

On the basis of these categories, the author formulated her proposal of evaluation protocol for postgraduate examinations, adding a “briefing” for evaluators (2006: 273). According to the author, more than 70% of the comments by the evaluators could have been classified according to the Effort Model proposed by Gile: of this percentage, about 70% were related to the production effort, about 15% to the listening effort and 14% to the coordination and memory effort (2006: 281).

At the end of her study, Soler Caamaño also formulated a proposal for a questionnaire of satisfaction, made of the following items: “el mensaje llega bien”; “los intérpretes demuestran un buen dominio del tema”; “la profesionalidad de los intérpretes es satisfactoria”; la expresión utilizada es agradable (voz, tono, ritmo); la terminología utilizada es adecuada”; “recomendaría/volvería a cotratar a los intérpretes”; “explique brevemente el porqué” (2006: 275). This questionnaire is almost the same as Doerflinger’s (2003), as the author herself recognized (Soler Caamaño 2006: 275).

PROPUESTA DE CUESTIONARIO DE SATISFACCIÓN					
		1	2	3	4
1.	El mensaje llega bien	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.	Los intérpretes demuestran un buen dominio del tema	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.	La profesionalidad de los intérpretes es satisfactoria	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.	La expresión utilizada es agradable (voz, tono, ritmo)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.	La terminología utilizada es la adecuada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
				<b>SÍ</b>	<b>NO</b>
6.	Recomendaría / volvería a contratar a los intérpretes que han hecho el servicio	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
7.	Explique brevemente el porqué de su respuesta anterior: .	.....			

Figure 6. “Propuesta – un cuestionario de satisfacción” (Soler Caamaño 2006: 275).

Nonetheless, at the end of her doctoral thesis the author maintains that a study on quality in conference interpreting should be carried out from a “holistic perspective” (Soler Caamaño 2006: 283).

According to Garzone (2003), it is possible that:

in the actual assessment of real instances of interpretation there might be interferences and interdependence between the different criteria separately submitted to, and evaluated by, respondents (Garzone 2003: 25).

Iglesias Fernández (2013: 59) proposes that “quality criteria do not seem to be processed separately, but holistically, in clusters of features”.

#### 4. CONCLUSION: THE NEED FOR A NEW PARADIGM OF QUALITY ASSESSMENT BASED ON GESTALT PERCEPTION OF SI

Considering the conclusions by Garzone (2003: 25) about the “interference” and “interdependence between the different criteria”, and by García Becerra (2013: 571) about the possibility that “insufficient formal aspects could eclipse remaining parameters in the evaluation mechanism”; and taking into account the proposal by Iglesias Fernández (2013: 59) about the possibility of a holistic processing of quality categories, and the proposal by Soler Caamaño (2006: 283) about the need for a “holistic perspective” in carrying out a study on quality assessment in conference interpreting, a theoretical paradigm for such a study has been created (De Gregoris, submitted). It seems that the need for a “holistic perspective” is also due to the fact that questionnaires used in the previous studies on quality evaluation (both expectations and assessment) have not changed substantially over time, as the chronological and contrastive table reported in Appendix shows. For this reason, the theoretical paradigm that was created inevitably has given rise to a proposal for a new questionnaire to elicit, in my case, subjects’ holistic perception of television broadcast simultaneous interpreting.

## AKNOWLEDGMENTS

I wish to express my deep gratitude to prof. Helena Lozano Miralles for editing and Mark Brady for reviewing this paper.

- Bühler H. (1986) "Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters", *Multilingua*, 5:4, pp. 231-235.
- Catana C. (2005) *Le qualità ben pronunciate: la dizione dell'interprete e la percezione dell'utente in simultanea*, tesi di laurea non pubblicata, SSLMIT, Università di Bologna, sede di Forlì.
- Chiaro D. & Nocella G. (2004) "Interpreters' Perception of Linguistic and Non- Linguistic Factors Affecting Quality: A Survey through the World Wide Web", *Meta*, 49:2, pp. 278-293.
- Collados Aís, Á.; Pradas Macías, E.; Stévaux, E. & García Becerra, O. (2007) (ed.) *La evaluación de la calidad en interpretación simultánea: parámetros de incidencia*, Granada, Comares.
- De Gregoris (submitted) "A proposal for a gestaltic assessment of voice quality in TV broadcast simultaneous interpreting: theoretical paradigm for a new questionnaire", in Proceedings of the International Conference ORALiTER, Forlì 3-4 December 2014, MonTI (Monografías de Traducción e Interpretación) Número especial - *La traducción de la oralidad* (2016). Alicante, Universidades de Alicante, Jaume I y Valencia.
- Doerflinger O. (2003) "Gestion de la qualité totale", in *La evaluación de la calidad en interpretación: investigación*. Ed by A. Collados, M. Fernández Sánchez, D. Gile, Granada, Comares, pp. 171-176.
- García Becerra O. (2013) *La incidencia de las primeras impresiones en la evaluación de la calidad de la interpretación simultánea: un estudio empírico*, tesis doctoral, Universidad de Granada.
- Garzone G. (2003) "Reliability of quality criteria evaluation in survey research", in *La evaluación de la calidad en interpretación: investigación*. Ed by A. Collados, M. Fernández Sánchez, D. Gile, Granada, Comares, 23-30.
- Gile D. (1990) "L'évaluation de la qualité de l'interprétation par les délégués: une étude de cas", *The Interpreter's Newsletter*, 3, pp. 66-71.
- Guardini P. (1995) *La traduzione simultanea del film: produzione e percezione. Un'analisi della prestazione dell'interprete e della risposta del pubblico*, tesi di laurea non pubblicata, SSLMIT, Università degli Studi di Trieste.
- Iglesias Fernández E. (2013) "Unpacking Delivery Criteria in Interpreting Quality Assessment", in *Assessment Issues in Language Translation and Interpreting*. Ed. by R. van Deemter, D. Tsagari, Frankfurt, Peter Lang, pp. 51-66.
- Kopczyński A. (1994) "Quality in conference interpreting: some pragmatic problems", in *Translation Studies: An Interdiscipline*. Ed by M. Snell-Hornby, F. Pöchhacker, K. Kaindl, Amsterdam/Philadelphia, John Benjamins, pp. 189-198.
- Kurz I. & Pöchhacker F. (1995) "Quality in TV Interpreting", *Traslatio-Nouvelles de la FIT- FIT Newsletter*, 14:3-4, pp. 350-358.
- Kurz I. (1993) "Conference interpretation: expectations of different user groups", *The Interpreter's Newsletter*, 5, pp. 13-21.
- Mack G. & Cattaruzza L. (1995) "User surveys in SI: a means of learning about quality and/or raising some reasonable doubts", in *Topics in Interpreting Research*. Ed. by J. Tommola, Turku, Centre for Translation and Interpreting, University of Turku, pp. 37-49.
- Marrone S. (1993) "Quality: a shared objective", *The Interpreter's Newsletter*, 5, pp. 35-41.
- Meak L. (1990) "Interprétation simultanée et congrès medical:

attentes et commentaires”, *The Interpreter’s Newsletter*, 3, pp. 8-13.

Moser P. (1995) “Survey on expectations of users of conference interpretation. Final report commissioned by AIIC”, file:///C:/Users/HP%206730S/Downloads/Full-report-in-PDF%20(2).pdf, consultato il 12/10/2014.

Ng B.C. (1992) “End-Users’ Subjective Reaction to the Performance of Students Interpreters”, *The Interpreter’s Newsletter*, Special Issue No. 1, pp. 35-41.

Palazzini Finetti, M (2000) *Interpretazione simultanea per un festival cinematografico: analisi linguistica e pragmatica dei risultati*, tesi di laurea non pubblicata, SSLMIT, Università degli Studi di Trieste.

Pöchhacker F. & Zwischenberger C. (2010) “Survey on quality and role: conference interpreters’ expectations and self-perceptions”, <http://aiic.net/page/3405/survey-on-quality-and-role-conference-interpreters-expectations-and-self-perceptions/lang/1>, consultato il 3/10/2014.

Russo M. (2005) “Simultaneous film interpreting and users’ feedback”, in *Interpreting*, 7:1, pp. 1-26.

Soler Caamaño, E. (2006) *La calidad en formación especializada en interpretación: Análisis de los criterios de evaluación de un jurado en un posgrado de interpretación de conferencia médica*, tesis doctoral, Universitat Pompeu Fabra.

Vuorikoski A.R. (1993) “Simultaneous interpretation-user experience and expectations”, in *Translation- The Vital Link. XIII FIT World Congress, Proceedings*. Ed. by C. Picken, Brighton, Institute of Translation and Interpreting, pp. 317-327.