# *Speech Audiometry in French-speaking Quebec*

# *Audiométrie vocale chez les francophones du Québec*

by • par

*Michel Picard, PhD*

Université de Montréal, Montreal, Quebec

## ABSTRACT

**A critical review of speech recognition tests currently used in Quebec with child and adult speakers of French is provided. Special attention is paid to tests used to determine speech recognition threshold (SRT) and their correspondence with pure-tone thresholds. The use and interpretation of SRT shifts in noise to identify possible sites-of-lesion are also reviewed as an alternative to traditional tests of percent word recognition score in quiet. Finally, the clinical utility of other testing procedures used in Quebec for French speakers such as the Speech Perception in Noise test for children is discussed.**

## ABRÉGÉ

**Examen critique des épreuves de reconnaissance de la parole en usage au Québec chez les jeunes et adultes francophones. On porte une attention particulière aux épreuves utilisées pour établir le seuil de reconnaissance de la parole (SRP) ainsi que sa correspondance avec les seuils pour les sons purs. On étudie également l'utilisation et l'interprétation des variations de SRP obtenues dans le bruit pour déterminer les sites de lésion éventuels. Cette approche est suggérée en remplacement des épreuves classiques de reconnaissance maximale de mots sans bruit. Enfin, on étudie l'utilité clinique d'autres méthodes d'évaluation employées au Québec auprès des francophones, comme l'épreuve de perception de la parole avec bruit.**

**KEY WORDS**

**speech audiometry • word recognition • speech perception • French**

The assessment of speech recognition is a central issue in audiological testing because speech is the auditory input most important to our daily communicative interactions (Olsen, 1991; Olsen & Matkin, 1991; Tillman & Olsen, 1973). Testing speech recognition is becoming increasingly important as we realise the shortcomings of pure-tone audiometry for determining a specific 'site-of-lesion' along the auditory pathway. Audiologists are increasingly aware that quantifying hearing for speech (as in speech audiometry) increases our understanding of the way human beings process complex stimuli (see Balota, 1994; Bernstein & Auer, 1996; Carpenter, Miyake, & Just, 1994; Lively, Pisoni, & Goldinger, 1994; Pichora-Fuller, 1996; Simpson, 1994 for contemporary reviews) and of the contribution of hearing loss to an auditory handicap (Noble, 1978). Speech audiometry can also help predict outcomes of rehabilitative strategies (Crowley & Nabelek, 1996; Gatehouse, 1994) and, consequently, is receiving increasing interest as a clinical intervention tool. Recent developments of the Speech Perception in Noise (SPIN; Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984; Elliott, 1995) and the Hearing in Noise Test (HINT; Nillson, Soli & Sullivan, 1994) are examples of this trend in developing tests to identify specific rehabilitative strategies.

Within this context, speech audiometry in French-speaking Quebec has developed significantly within the last few years. Back in the sixties and early seventies, there were no recorded materials available from France to assess speech audiometry of French speakers and speakers of English were typically tested with materials developed in the United States (as they still are today). Recordings of the *Test phonétique* by Lafon (1958; a test developed in France for determining the contribution of recruitment) were not available in Quebec because the assumption was that word lists (to be presented live voice) — and not some standard recordings of these lists — represented the real test items. Moreover, no attempt was made to validate this test with large samples of individuals typically seen in the audiology clinic. There were similar problems with all materials developed by Fournier (1951). In other words, French tests were presented as valid instruments while, in fact, there was no guarantee of validity (i.e., reasonable statistical data). Thus, they represent typical examples of 'face' or 'faith' validity (Guilford, 1954). This situation has been the primary motivation for a collective endeavour in French-speaking Quebec to refine protocols in speech audiometry. As a consequence, a small number of valid instruments are now available and several others are in development (Lebel & Picard, 1995, 1997; Picard, 1984; Picard, Banville, Barbarosie, & Manolache, in press; Tremblay, Picard, Barbarosie, & Banville, 1991). In fact, speech audiometry in French-speaking Quebec may deserve special consideration not only because of professional services to speakers of French, but also because of the difference in clinical protocols used in French-speaking Quebec and those used in the United States (Berger, 1978; Bess & Humes, 1990; Goetzinger, 1978; Hopkinson, 1978; Kruger & Mazor, 1987; Martin, 1978, 1997; Newby, 1979; Olsen & Matkin, 1991; Rupp, 1980; Schill, 1985; Silman & Silverman, 1991; Stockdell, 1980; Tillman & Olsen, 1973; Wilson & Margolis, 1983). Speech audiometry protocols developed in French-speaking Quebec have used standardised recordings of materials to avoid the limitations imposed by increased variability in test scores as a result of live voice administration (Penrod, 1994). This is another divergence from preferred audiometric practices in the United States (Martin & Morris, 1989; Martin & Sides, 1985). Thus, efforts made in French-speaking Quebec to develop innovative and standardised tests adhering to goals of test accuracy that extend beyond the alleged rapidity, flexibility, and ease of administration associated with monitored live voice presentations (Penrod, 1994) may be of paramount importance in speech audiometry.

Speech testing protocols in French-speaking Quebec, which are the main focus of this article, have been developed from, roughly, two theoretical lines. The Articulation Index (AI) theory (Fletcher, 1929; French & Steinberg, 1947; Pavlovic, 1993; Studebaker & Sherbacoe, 1993), a quantification scheme proposed by the telephone industry to account for audibility of speech cues, served as the basis for two measures. Speech recognition threshold (SRT) was proposed to represent the minimal amount of information a person using the telephone would require to understand conversational speech (Carhart, 1946a; Hudghson & Thompson, 1942). Percent word recognition score (PWRS) represented the maximum number of bits of information (or speech cues) that could be perceived in the speech signal (i.e., speech recognition as a test of 'auditory acuity' as Carhart, 1946b, put it). For clinical purposes in this particular framework, the human listener was, of course, considered the transmission system under study (Fletcher, 1953) instead of a telecommunication system like the telephone, as originally conceived by the engineers at Bell Telephone Laboratories (Fletcher, 1929; French & Steinberg, 1947). Because human listeners were involved, a proficiency factor (p) (i.e., an exponent to account for the linguistic competence and experience of the listener; see Pavlovic, 1993; Studebaker and Sherbacoe, 1993, for contemporary reviews) had to be incorporated into the AI mathemetical expression. At the time audiologists were more concerned with the basic skills of encoding sensory input and they were convinced that speech tests could be developed as almost pure measures of performance and that the contributions of linguistic and/or cognitive factors were negligible. Thus, the proficiency factor was defined as a constant ($p=1$; Studebaker & Sherbacoe, 1993). One serious limitation with this approach, of course, is that human beings depart from optimal transmission systems for a variety of reasons (Gordon-Salant, 1986; Picard et al., in press). As well, there is considerable evidence that speech stimuli require more complex signal processing strategies than pure tone stimuli (see Balota, 1994; Bernstein & Auer, 1996; Carpenter et al., 1994; Lively et al., 1994; Pichora-Fuller, 1996; Simpson, 1994, for contemporary reviews).

The second trend in the development of speech audiometry in Quebec was derived from the Netherland school, with Plomp (1978, 1986) as the most visible proponent. Plomp proposed a model of speech recognition in noise that differs from listening in quiet in that speech recognition in noise is governed by factors related to frequency selectivity (Festen & Plomp, 1983) and temporal resolution (Dreschler & Plomp, 1985). By contrast, speech recognition in quiet could be essentially attributable to audibility factors. This framework has been used by researchers at the *Université de Montréal* to develop speech recognition tests in noise that would complement the traditional SRT test and PWRS in quiet. The goal of the present article will now be to review these various test protocols and discuss their clinical utility.

## A Short History of Speech Recognition Test Development in French-speaking Quebec

Speech materials and protocols to assess hearing for speech in Quebec were developed as a result of intensive research activities in the late forties to adapt research materials developed at Bell Telephone Laboratories for the clinical evaluation of adults with hearing impairments (Carhart, 1946a, 1946b; Egan, 1944; 1948; Hirsh, Davis, Silverman, Reynolds, Eldert, & Benson, 1952; Hudghson & Thompson, 1942). What resulted was a test protocol of four recorded lists of 42 familiar polysyllables used to determine SRT and eight lists of 50 phonetically balanced (PB) monosyllables to assess PWRS. This protocol was suggested by Benfante et al. (1966) to replace the word lists in French that were used at the time (i.e., the material by Fournier, 1951). However, the clinical efficacy of materials developed by Benfante et al. was only evaluated much later by Picard (1984) at the *Université de Montréal*. Various studies conducted during this period, mostly with noise-exposed workers, led to a revised set of 11 highly recognisable words for SRT measurement and four (of the eight original) PB-word lists to determine PWRS. This material was used by Tremblay et al. (1991) to determine speech recognition threshold shift in noise (SRTSN) and provided an alternative to PWRS and a more refined site-of-lesion test. Many authors have called for an instrument that would tap a broader spectrum of the numerous listening skills involved in speech recognition (Balota, 1994; Bernstein, 1996; Carpenter et al., 1994; Gierut & Pisoni, 1988; Kent, 1992; Lemme & Hedberg, 1988; Lively et al., 1994; Pichora-Fuller, 1996; Pisoni & Luce, 1987; Simpson, 1994; Strange, 1986). Consistent with this expressed need, Lebel and Picard (1995) proposed a protocol to determine SRT in children that was devoted to the assessment of factors beyond audibility contributing to SRT (Lebel & Picard, 1997). Further, Hébert and Picard (1979) described a PWRS for children using picture identification. This test was developed in a manner similar to the Word Intelligibility Through Picture Identification (WIPI) test by Ross and Lerman (1970). The material was further expanded by Lefebvre (1991) to incorporate words of high- and low- predictability (HP and LP) in sentences as a preliminary version of a SPIN test for children. Recordings of all these materials are available on cassette from the *Université de Montréal* and are currently being transferred onto compact disks (available as of April, 1998).

## Testing SRT in Adults

Testing SRT in French-speaking adults in Quebec is often done using Picard's randomisations (1984) of the 11 most recognisable bisyllables identified by Bernatchez and Toupin-Rochon (1977). Essentially four lists of 22 words, this material is used to determine both SRT in quiet (Picard, 1984) and SRT shift in noise (Tremblay et al., 1991). These materials are presented in Appendix A.

*SRT in quiet.* Picard (1984) proposed that the word lists described in Appendix A be presented at sensation levels in blocks of five items to construct the steep portion of the psychometric function that would include SRT (i.e., a performance-intensity function expressing percent correct scores as a function of presentation levels as a means to bracket the 50% performance level). By definition, this level is set to correspond to SRT. However, because the method uses five words per presentation level, it produces scores changing in 20% step sizes (i.e., 20, 40, 60%, and so on), and the 50% performance level is then interpolated from the psychometric function.

Using this method, the initial testing will be typically initiated at 10 dB above the pure-tone average (PTA; i.e., the Fletcher index, 1950, corresponding to the average of the two smallest values of hearing loss at 0.5, 1, and 2 kHz). The next presentation level will depend on the observed performance. If below 50% correct identification, the next presentation level will be increased in 10-dB steps up to the point it will exceed the 50% performance level. Conversely, if the initial observed performance is above 50% correct identification, the next presentation level will be decreased in 10-dB steps as long as performance stays above 50% correct identification. Eventually, the run will be reversed in 5-dB ascents whenever performance falls below 50%. The SRT can thus be interpolated with a reasonably small number of blocks (usually three) providing some reasonable compromise between speed of execution and test-retest reliability. On that particular issue, Plomp and Mimpen (1979) have already indicated that determining the SRT in 2-dB steps with sets of only five words provides a test-retest difference of only 1.3 dB. This compares favourably with the 0.9 dB test-retest difference found when sets of 10 to 13 items are used. A 5 dB step size method would slightly inflate errors, although not to the point of making it larger than the error of measurement associated with pure-tone audiometry (ANSI, 1978). In fact, the difference between the 2 and the 5 dB is clinically insignificant; the 5-dB method being responsible for a 1.83 dB SRT elevation, on average, compared to the 2-dB method (Chaiklin & Ventry, 1964). Therefore, the 5x5 method proposed by Picard (1984; i.e., blocks of five words in 5-dB ascents) results in a marginal loss of test-retest reliability compared to longer test protocols (Plomp & Mimpen, 1979).

Recent studies by Picard et al. (in press) revealed two particular features of the test materials and protocols developed in 1984. First, SRTs measured using this protocol were systematically smaller than the Fletcher index, suggesting that the material developed in 1984 may be more perceptible than those used in the United States (eg., the W-1 and W-2 test words by Hirsch et al., 1952, which serve as a reference point for audiometer calibration, ANSI, 1989). The overall difference between SRT and PTA corresponding to SRT minus PTA $_{\text{Fletcher}}$ (or speech-to-pure-tone difference, SPD) was found to be -2.16 dB in 807 observations obtained from noise-exposed workers. Picard et al. (in press) also suggested that the 1984 test material influenced subject responses in 28.5% of the 807 observations collected, to produce SRTs far more perceptible than the Fletcher index. More specifically, six subsamples of participants with various degrees of SRTSN showed, on average, SRTs in quiet 6.74 to 15.96 dB smaller than the Fletcher index. This surprising outcome was attributed to Bruce's findings (1956) that these listeners were aware, to some degree, that the number of word alternatives was reduced during SRT measurement (the speech material being limited to a set of only eleven words).

Concluding that SRTs smaller than the Fletcher index resulted from a sophisticated guessing strategy used by a small group of, apparently, linguistically proficient listeners was motivated by the following two considerations: (a) the presence of extremely high correlations between SRT and the Fletcher index in the six subsamples where this particular type of context may have been detected (Pearson's *r* ranged from .85 to .96); and, (b) the closer correspondence between SPD and SRT shifts in noise in, presumably, more sophisticated listeners (n=230), compared to listeners not showing the same effect (n=577), which resulted in a significant increase in negative coefficients of correlation ($p < 0.01$). Interestingly, Tremblay et al. (1991) attributed SRT shifts in noise primarily to sensory factors increasing the demand for signal processing on linguistic and cognitive resources. As a whole, these two indications do not support the contention that a SRT smaller than the Fletcher index by amounts ranging from 7 to 16 dB would result from pseudo-hypoacusis as would often be suspected by clinical audiologists (Berger, 1978; Bess, 1983, 1988; Cooper, 1980; Kruger & Mazor, 1987; Noble, 1978; Olsen & Matkin, 1991; Tillman & Olsen, 1973; Ventry, 1976).

The changes in response criteria identified by Picard et al. (in press) in a large number of observations have led to a revised categorisation of the correspondence between SRT in quiet and the Fletcher index. Table 1 summarises confidence limits when predicting SRT from pure-tone thresholds. Examination of this table reveals a ±5 dB correspondence between SRT and the Fletcher index for the vast majority of individuals (95%) during the first determination of SRT in quiet. However, participants who become 'sophisticated' guessers (realising that they are being offered a reduced number of alternatives to choose from) may depart by as much as 10 dB from the Fletcher index when listening to barely audible speech (SRT smaller than the Fletcher index). In addition, some individuals fall outside both of these categories. One profile will correspond to SPDs exceeding the upper confidence limits shown in Table 1. This particular behaviour would often be interpreted as a sign of limited linguistic competence (including nonnative listeners), according to Nabelek and Nabelek (1994), Takata and Nabelek (1990), and Borchgrevink (1986). Alternatively, this phenomenon could result from language pathology (Kruger & Mazor, 1987; Silman & Silverman, 1991) such as aphasia (Caplan & Utman, 1994). Conversely, when speech is far more perceptible than pure tones (resulting in SPDs smaller than the one might expect from sophisticated guessers), then — and only then — would pseudo-hypoacusis be considered.

**Table 1. Confidence limits to the correspondence between SRT in quiet and the Fletcher index when using the speech material and protocol proposed by Picard (1984).**

| Clinical category | N | SEM[1] (dB) | Confidence limits to predict SRT from Fletcher Index (SPD in dB) | |
|---|---|---|---|---|
| | | | 10-90th percentile | 2.5-97.5th percentile |
| Normal correspondence between SRT and PTA | 577 | 2.858 | -3.67 to 3.64 | -5.61 to 5.58 |
| Sophisticated guessing | 230 | 1.719 | 9.73 to -5.33 | -10.9 to -4.16 |

[1] Standard of measurement of SRT; that is, the extent of dispersion of error components in SRT when predicted from the Fletcher index. SEM was derived from Hoyt's analysis-of-variance approach to reliability (Guilford, 1954).

Quite clearly, listeners may be detecting linguistic contexts during SRT measurement when severe restrictions are imposed on the testing vocabulary (Picard et al., in press) and this may have a significant impact on speech recognition, SRT in quiet in particular. This may suggest that with appropriate experimental controls the SRT-PTA relationship can be made less sensitive to the confounding influences of linguistic factors. Awareness of a reduced number of alternatives to choose from when severe restrictions are imposed on testing vocabulary would indeed bias participant responses towards a maximal contribution of language proficiency. Thus, when participants anticipate results, this particular type of response bias would likely help fulfil the requirement of a proficiency factor (p) equal to unity in Articulation theory (Studebaker & Sherbacoe, 1993). In turn, *p* as a constant of one would result in a SRT-PTA relationship which is essentially determined by audibility factors. Therefore, determining SRT with a restricted vocabulary seems to represent a legitimate means to identify contributions to speech recognition in conditions of high stimulus uncertainty that may have otherwise confounded results (resulting in speech far more perceptible than pure tones) with pseudo-hypoacusis (Bess, 1983, 1988; Cooper, 1980; Ventry, 1976). Clearly, this is a gain in test accuracy and significantly improves the clinical utility of speech protocols. For instance, finding that someone is shifting to a sophisticated guessing strategy may serve as a predictor variable of success in any rehabilitative program or activity offered to him that would capitalise on the efficient use of language context effects.

*SRT shift in noise.* Tremblay et al. (1991) proposed that the measurement of SRT shifts in noise could represent the distortion (class D) component identified by Plomp (1978, 1986) in SRT in noise. They recommended that SRT in quiet be repeated over a background noise (broadband speech spectrum noise) simultaneously mixed with speech in the same ear. The signal-to-noise ratio (SNR) was maintained at 0 dB for the duration of the test. The method of measurement was identical to that mentioned above used to determine SRT in quiet, except that runs were started at 5 dB above SRT in quiet and pursued in 5-dB ascents up until the 50% performance level was exceeded for the first time. When this level of performance was not reached easily, the procedure was prolonged, though never beyond twelve consecutive increases in presentation level (65 dB SL) to avoid risks of uncomfortable listening levels. After completion of the procedure, SRT shift in noise was expressed as the difference between the two SRTs (SRT in noise minus SRT in quiet). When SRT in noise could not be determined, it was indicated as a value in excess of permissible clinical limits in a manner similar to pure-tone thresholds when exceeding the range of measuring equipment.

Studying noise-exposed workers, Tremblay et al. (1991) identified six clinical categories of SRT shifts in noise. These categories are summarised in Table 2 along with their recent reinterpretation by Picard et al. (in press). In essence, Table 2 shows SRT shifts in noise determined by disruption (or failure) of lexical access in noise. Difficulties accessing the lexicon would in turn be associated with Plomp's class D component of SRT in noise (1978, 1986) acting as an intermediate variable to add some extra load on linguistic and/or cognitive resources. Relaxation in subject response criteria as a result of interfering noise was also found to plague performance of groups with level

1 and 2 SRT shifts in noise in a manner similar to those of elderly subjects (Gordon-Salant, 1986). In addition, a nonmeasurable SRT shift in noise might indicate some severe sensory (cochlear or retrocochlear) damage, or cognitive involvement. Here, cognitive involvement refers to auditory dysfunctions possibly resulting from limitations of language competence (Nabelek & Nabelek, 1994), language pathology (Kruger & Mazor, 1987; Silman & Silverman, 1991), or overloading of working memory resources resulting from their momentary reallocation to support speech recognition in particularly demanding listening conditions (Pichora-Fuller, Schneider, & Daneman, 1995).

The accuracy of Tremblay's classification was found to be extremely high (90.8%). Moreover, with 55.76% of all scores departing from normal (450/807), this test showed greater sensitivity to deterioration of signal processing by the ear than traditional measures of PWRS. In contrast, it is worth mentioning that Carhart (1965) reported that only 39.4% of his clinical sample fell below the 90% cut-off point to separate supposedly 'normal' from 'abnormal' speech recognition scores (Goetzinger, 1978). Thus, there is little doubt that the test proposed by Tremblay et al. (1991) is both sensitive and accurate to differentiate speech understanding skills in noise. That would include conditions when lexical access suffers from restrictions imposed on sensory information (as a result, for instance, of Plomp's class D component of SRT in noise) or from relaxation in subject response criteria. Complementary evidence arises from data provided by Plomp (1994) indicating that individuals with Meniere's disease and those with presbyacusis are especially vulnerable to deterioration of SRT in noise.

This level of understanding of SRT shifts in noise, however, may fall short of ensuring a level of clinical performance that would optimally differentiate sites-of-lesion. In particular, the study by Tremblay et al. (1991) is probably not comprehensive enough to fulfil this particular requirement and SRTSN was determined only with noise-exposed workers. Nevertheless, the accuracy of the clinical categories shown in Table 2 clearly indicates that the instrument is capable of characterising group differences. SRT shifts in noise may prove vital to the formulation of reasonable hypotheses about the underlying locus of a given auditory dysfunction even if it may lack the necessary specificity to differentiate specific contributions of sensory and cognitive factors.

Furthermore, the particular sensitivity of SRT shifts in noise can be used to enhance the predictive power of a test protocol. For instance, as suggested by Turner (1988, 1991), the test may be combined in series with a very specific and noncorrelated (or only partially correlated) instrument to increase performance. Given that the accuracy of the predictions made with a particular test protocol depends upon the accuracy of individual test performance data and the proportion of shared variance across tests (i.e., the tendency of individual tests to identify the same patients as positive or negative, Turner, 1988), this effort would certainly represent a gain in test protocol efficacy. Applied to the particular context of speech audiometry, a protocol that specifically aims at differentiating cochlear from retrocochlear site-of-lesion could begin with SRT shift in noise, and be followed by traditional PWRS, depending on the first test result. When a nonmeasurable SRT shift in noise is identified, PWRS could be introduced to differentiate individuals with extremely

**Table 2. Classification of speech recognition threshold shifts in noise ($SRT_{sn}$) according to Picard's re-interpretation of Tremblay et al. clinical categories ($SRT_{sn}$ range from the original study).**

| Clinical category | N of observations | $SRT_{sn}$ range (dB) | Interpretation |
|---|---|---|---|
| no $SRT_{sn}$ | 357 | 0/10 | Normal limits. Variations in $SRT_{sn}$ due to occasional failure of lexical access in noise; intermediate variable associated with this failure: Plomp's class D component of SRT in noise; |
| $SRT_{sn}$ level 1 | 166 | >10/15 | Mild $SRT_{sn}$ due to occasional failure of lexical access in noise; intermediate variables associated with this failure: Plomp's class D component of SRT in noise and relaxation in subject response criteria as a result of speech interference by noise; |
| $SRT_{sn}$ level 2 | 76 | >15/20 | Moderate $SRT_{sn}$ due to occasional failure of lexical access in noise; intermediate variable associated with this failure: Plomp's class D component of SRT in noise and relaxation in subject response criteria as a result of speech interference by noise; |
| $SRT_{sn}$ level 3 | 80 | >20/30 | Severe SRTsn due to failure of lexical access in noise; intermediate variable associated with this failure: Plomp's class D component of SRT in noise; |
| $SRT_{sn}$ level 4 | 50 | >30/64 | Extreme $SRT_{sn}$ due to general failure of lexical access in noise; intermediate variable associated with this failure: Plomp's class D component of SRT in noise; |
| not measurable | 78 | >64 | Totally impaired lexical access or giving-up in noise; intermediate variable(s) associated with this behaviour: undetermined. |

low scores indicative of VIII nerve involvement (scores from 15 to 58% according to Turner, Shepard, & Frazer, 1984). In this framework, higher scores might indicate some sensory (cochlear) or cognitive involvement.

## Measurement of PWRS in Adults

Picard (1984) also identified four lists of monosyllabic words that produced a similar level of performance in noise-exposed workers when presented at 32 dB above SRT in quiet. These lists are presented in Appendix 2. Studying the distribution of scores in 48 workers with noise-induced hearing loss, the author proposed the five clinical categories summarised in Table 3. However, given the binomial distribution and variability of PWRS (Thorton & Raffin, 1978), Picard (1984) discussed the severe limitations on the clinical usefulness of these categories. For instance, he found that percent scores of Category 1 (between 92 and 100%) had up to a 36% chance of being mistakenly classified as scores of Category 2 (82-90% range). Similarly, for other categories: scores of Category 2 were found to be accurate only between 45 and 55% of the time; scores of Category 3, only between 41 and 43%; and scores of Category 4, only between 42 and 50%. In other words, findings by Picard (1984) warned against the indiscriminate use of PWRS, as the low predictive power of this test inevitably curtails its clinical usefulness (Carhart, 1965; Keith, 1988).

**Table 3. Guidelines for evaluating maximum percent word recognition scores (PWRS).**

| Range of Scores | Clinical Category |
|---|---|
| Cat. 1, 92-100 | Normal limits |
| Cat. 2, 82-90 | Slight listening difficulties as a result of cochlear lesion |
| Cat. 3, 72-80 | Moderate listening difficulties as a result of cochlear lesion |
| Cat. 4, 58-70 | Poor speech recognition as a result of cochlear lesion |
| Cat. 5, ≤ 56 | Very poor speech recognition as a result of cochlear or retrocochlear lesion |

## Measurement of SRT in Children

In the early nineties, Lebel and Picard (1995) developed a speech protocol to determine SRT in school-age children. The material incorporated 18 familiar bisyllables in picture form. The words were randomised to create the five lists of 18 items presented in Appendix 3. Each list was accompanied by a particular randomisation of 36 coloured pictures presented in groups of six on six separate response sheets. Children were asked to identify only three words per response sheet to keep chance responding to 16.6% on the first, 20% on the second, and 25% on the third trial.

SRT was determined using essentially the same procedure as proposed by Picard (1984) except that a reduced set of words was used per presentation level. Specifically, the number of words was reduced from five to three, to speed up the test which requires the potentially difficult task of tracking auditory stimuli at the threshold of audibility. Using only three words per level, as opposed to 10 to 13, may explain a loss of 0.7 dB in test-retest reliability according to Plomp and Mimpen (1979). It was believed that this potential loss of reliability was a reasonable trade-off for a shortened — and consequently, a potentially more easily administered — test protocol with children. Findings in 12 five-year-old children with normal hearing revealed a psychometric function with the same steep growth as is usually obtained from adult listeners and a reasonable correspondence between SRT and average pure-tone thresholds at 0.5, 1, and 2 kHz (±6.6 dB).

In a second study, Lebel and Picard (1997) compared the performance of three groups of children, controlling for response format as a means to verify the particular influence of lexical access on SRT. More specifically, the study was conducted with groups of 24 children with normal hearing aged between six and 11 years. Quite surprisingly, results showed a significant benefit of a closed-set response format over straight repetition of words even in the oldest groups. SRTs were, on average, 5 dB more perceptible using a response format that restricted the number of possible word candidates. This finding strongly suggested language context effects on the SRT task. The authors recommended that a closed-set response format be used, rather than the more frequently used method of repeating words (Martin & Morris, 1989; Martin & Sides, 1985). This recommendation was seen as a rather simple way to reduce the confounding influence of language context effects in a clinical test protocol where the correspondence of SRT and PTA is of particular importance as a measure of internal consistency (or concurrent validity) of the two forms of tests.

## Measurement of PWRS in Children

The determination of maximum speech recognition score also attracted the attention of researchers in French-speaking Quebec. In particular, Hébert and Picard (1979) prepared a set of four lists of 20 monosyllabic words familiar to children aged five and older (see Appendix D). Children were required to match each word with a picture using a closed-set response format in a manner similar to the WIPI test proposed by Ross and Lerman (1970). The closed-set response format was comprised of six alternatives, including two placebos, which limited chance responding to 16.6%. Target C-V-C words differed from one list to the other by only a few distinctive features (for instance, 'pain, bain, main, faim'). There were 14 words contrasted on the initial consonant, one on the final consonant and five on the vowel. However, because of the limited number of items, the clinical utility of this instrument to predict a particular site-of-lesion was uncertain. A 20-item test will inevitably carry much less predictive power than a longer test (such as, one with 50 items, which is usual in speech audiometry). Again, this is mostly attributed to the binomial distribution of scores (Thornton & Raffin, 1978) which compromises criterion validity (Turner et al., 1984). Given the stated limitations, this test has been used mostly for research purposes in audiology. An unexpected application of this test, however, was made by clinicians in speech-language pathology interested in assessing phonological development in children. It elucidated differences between errors in the perception and production of speech sounds. The Hébert-Picard test (1979) has also been proposed as a short test to determine PWRS and classify subjects globally into clinical groups.

Recently, Lefebvre (1991) redefined the goal of the Hébert-Picard test by proposing a SPIN test (Kalikow, Stevens, & Elliott, 1977) based on two scramblings of List 3 from Hébert and Picard (1979). This material is presented in Appendix E. Similar to the original SPIN test in many respects, the French version proposed by Lefebvre used broadband speech spectrum noise as a masker instead of speech babble (12 talkers simultaneously reading continuous text as in the original SPIN test). SNR was set to +10 dB based on the original work by Kalikow et al. (1977). Of particular importance, each list was made of 10 sentences where the final word was predictable from the sentence context (high-predictability word), and 10 sentences where the final word was not predictable (low-predictability word).

Table 4 summarises the main findings obtained with this test in 12 five- to six-year-old kindergarten children with normal hearing and no history of recurrent otitis media. Examination of Table 4 shows global scores ranging from 27.9 to 37.5%, depending on the list. Furthermore, subtests with high- and low-predictability sentences differed by as much as 19.2% (HP minus LP score). A difference of this magnitude between HP and LP scores is in agreement with Nittrouer and Boothroyd (1990) for children of the same age. However, global scores of 27.9 to 37.5% are much lower than the scores of 54% obtained for words in isolation at +3 dB SNR and 65% at 0 dB SNR for combined HP and LP sentences (Nittrouer & Boothroyd, 1990). Studies by both Lefebvre and Nittrouer and Boothroyd used an open-set response format (children repeated what they heard). However, it is possible that the French paediatric version of Lefebvre (1991) presented a more difficult listening situation, resulting in the use of more relaxed response criteria for children (Craig, Kim, Pecyna Rhyner, & Bowen Chirillo, 1993). This possibility is in agreement with the findings of Lively et al. (1994) indicating word repetition especially susceptible to response biases and guessing strategies. In particular, the use of the five- to eight-word sentences of Lefebvre (1991) instead of only four monosyllables in the study of sentence-level context effects by Nittrouer and Boothroyd (1990) may be responsible for a different decision strategy used by children. Moreover, the current French version of the SPIN test involves only two lists of 20 sentences. This is in contrast to 10 lists of 50 sentences in the original SPIN test (Kalikow et al., 1977), eight lists of 50 sentences in the SPIN-R (Bilger et al., 1984), and 80 four-word sentences in the Nittrouer and Boothroyd (1990) study, each serving as stimulus and as context for the other words. Clearly, there are large differences in test size between instruments in French and in English.

**Table 4. Means (*M*) and Standard Deviation (*SD*) for speech recognition scores obtained by kindergarten children on the French version of the SPIN test (Lefebvre, 1991).**

| Scoring method | *M* | *SD* |
|---|---|---|
| List 3A | | |
| Total score | 27.91 | 10.69 |
| High-predictability | 37.50 | 11.63 |
| Low-predictability | 18.33 | 12.80 |
| List 3B | | |
| Total score | 37.50 | 15.34 |
| High-predictability | 45.00 | 18.93 |
| Low-predictability | 30.00 | 19.58 |

## Future Directions in Speech Audiometry

This review has illustrated that speech audiometry in French-speaking Quebec is a work in progress to adapt and develop tests that would support traditional clinical decisions. Our efforts have been taking place primarily at the Université de Montréal and more recently at the University of Ottawa. Some efforts have also been initiated to export materials developed in French-speaking Quebec to French-speaking groups outside of the province. The University of Ottawa is also working on a French version of the HINT test by Nillson et al. (1994).

Speech audiometry in French-speaking Quebec has benefited from the special attention paid to the correspondence between SRT and PTA. Furthermore, innovative instruments like SRTSN and the SPIN test offer potentially unique opportunities to understand the intricate auditory processes involved in the organisation of sensory information (corresponding to the audibility factor) into some efficient and meaningful streams of linguistic information. This may be particularly important given the increasing evidence of the influence of linguistic and cognitive variables on tests such as the SRT and PWRS. For instance, the SRTSN reveals that subject responses are determined by relaxation of response criteria when a masker exacerbates the insult of cochlear damage in noise-exposed workers (Picard et al., in press). Conversely, when determining SRT in quiet, the search for plausible word candidates in the lexicon may be facilitated in listeners if they become aware that the number of alternatives from which to choose is reduced when severe restrictions are imposed on the testing vocabulary. Similarly, the use of sentence-level context effects by kindergarten children as revealed by the SPIN test is quite compelling evidence of these types of cognitive-linguistic effects on speech tests currently used in clinical audiology.

Clearly, underlying assumptions of speech audiometry in French-speaking Quebec have been significantly revised, and the current test instruments provide a much better assessment of the way human beings process speech. Not only does speech audiometry provide useful information of the relative contribution of sensory and cognitive factors in speech understanding, but as indicated by Crowley & Nabelek (1996), by Plomp (1994), and by Gatehouse (1994), it also provides for the evaluation of the efficacy of such audiological treatments as hearing aids.

After 50 years of research in speech audiometry and reassessment of its clinical usefulness, the particular contexts (or aspects) of speech recognition that are important to measure continue to be revisited. Recognition of speech sounds by the ear does not necessarily mean that syllables, words (or even larger linguistic segments) will also be recognised. So, one may wonder what clinicians in audiology are really interested in when measuring speech recognition? We need to determine what level of linguistic input we are interested in assessing. Instruments such as the HINT and the SPIN clearly are directed towards the assessment of word recognition and its influence by sentence-level contextual effects. Should other types of language contexts (for instance, those ones due to restrictions on the testing vocabulary) also be taken into account in the construction of more comprehensive profiles of patient's listening skills? Alternatively, we might also design a test that maximises the potential effects of language-context, and this in turn may maximise speech recognition. In this way, failure will most likely be due to audibility factors. That is, such tests might allow for the differentiation of cognitive-linguistic factors from those more simply due to the hearing loss. One of the fundamental questions facing audiologists is what exact level (or levels) of speech recognition should be assessed. We certainly know at this time that speech audiometry is addressing issues far more complex than originally imagined.

## Acknowledgements

Thanks to David McFarland for his thoughtful and instructive comments on earlier versions of the paper.

***Please address all correspondence to:*** *Michel Picard, École d'orthophonie et d'audiologie, Université de Montréal, CP 6128, succ. Centre-ville, Montréal, QC, H3C 3J7.*

## References

American National Standards Institute. (1978). *American national standard methods for manual pure-tone threshold audiometry. S3.21*. New York: ANSI.

American National Standards Institute. (1989). *Specifications for audiometers. S3.6*. New York: ANSI.

Balota, D. A. (1994). Visual word recognition. In M. A. Gernsbacker (Ed.), *Handbook of psycholinguistics* (pp. 303-358). New York: Academic Press.

Benfante, H., Charbonneau, R., Arseneault, A., Zinger, A., Marti, A., & Champoux, N. (1966). *Audiométrie vocale*. Montreal: Hôpital Maisonneuve.

Berger, K. (1978). Speech audiometry. In D. Rose (Ed.), *Audiological assessment* (2nd ed., pp. 227-260). New York: Holt, Rinehart, & Winston.

Bernatchez, C., & Toupin-Rochon, C. (1977). *Vérification de l'homogénéité des dissyllabes de Benfante*. Unpublished master's thesis, Université de Montréal, Montreal, Quebec.

Berstein, L. E., & Auer, E. T. (1996). Word recognition in speechreading. In D. G. Stork & M. E Hennecke (Eds.), *Speechreading by humans and machines*. NATO ASI series (pp. 17-26). Berlin: Springer Verlag.

Bess, F. (1983). Clinical assessment of speech recognition. In D. F. Konkle & W.F. Rintelmann (Eds.), *Principles of speech audiometry*. Baltimore: University Park Press.

Bess, F. (1988). Basic hearing measurement. In N. J. Lass, L. V. Reynolds, J. L. Northern, & D. E. Yoder (Eds.), *Handbook of speech-language pathology and audiology*. Toronto: Decker.

Bess, F.H., & Humes, L.E. (1990). *Audiology: The fundamentals*. Baltimore: Williams & Wilkins.

Bilger, C., Nuetzel, M. J., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27*, 32-48.

Borchgrevink, H. M. (1986). Reference coherence in speech perception: Consequences for native and second language speech comprehension in noise. In R. J. Salvi, D. Henderson, R. P. Hamernick, & V. Colletti (Eds.), *Basic and applied aspects of noise-induced hearing loss* (pp. 357-367). New York: Plenum.

Bruce, D. J. (1956). Effects of context upon intelligibility of heard speech. In C. Cherry (Ed.), *Information theory: Third London symposium* (pp. 123-131). London: Butterworth.

Caplan, D., & Utman, J. A. (1994). Selective acoustic phonetic impairment and lexical access in an aphasic patient. *Journal of the Acoustical Society of America, 95*, 512-517.

Carhart, R. B. (1946a). Speech reception in relation to pattern of pure tone loss. *Journal of Speech and Hearing Disorders, 11*, 97-108.

Carhart, R. B. (1946b). Monitored live-voice as a test of auditory acuity. *Journal of the Acoustical Society of America, 17*, 339-349.

Carhart, R. B. (1965). Problems in the measurement of speech discrimination. *Archives of Otolaryngology, 82*, 253-260.

Carpenter, P. A., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M. A. Gernsbacker (Ed.), *Handbook of psycholinguistics* (pp. 1075-1122). New York: Academic Press.

Chaiklin, J. B., & Ventry, I.R. (1964). Spondee threshold measurement: A comparison of 2- and 5-dB methods. *Journal of Speech and Hearing Disorders, 29*, 47-59.

Cooper, J. C. (1980). Speech protocols in the assessment of functional hearing loss. In R. R. Rupp & K. G. Stockdell (Eds.), *Speech protocols in audiology* (pp. 227-252). New York: Grune & Stratton.

Craig, D. H., Kim, B. W., Pecyna Rhyner, P. M., & Bowen Chirillo, T. K. (1993). Effects of word predictability, child development and aging on time-gated speech recognition performance. *Journal of Speech and Hearing Research, 36*, 832-841.

Crandell, C. C. (1991). Individual differences in speech recognition ability. *Ear & Hearing,12* (Suppl.), 100S-108S.

Crowley, H. J., & Nabelek, I. V. (1996). Estimation of client-assessed hearing aid performance based upon unaided variables. *Journal of Speech and Hearing Research, 39*, 19-27.

Davis, H., & Silverman, R. S. (1978). *Hearing and deafness* (4th ed.). New York: Holt, Rinehart & Winston.

Dreschler, W. A., & Plomp, R. (1985). Relation between psychophysical data and speech perception for hearing-impaired subjects. Part II. *Journal of the Acoustical Society of America, 78*, 1261-1270.

Egan, J. P. (1944). *Articulation testing II.* OSRD report 3802 (PB22848). Washington, DC: Office of Scientific Research and Development.

Egan, J. P. (1948). Articulation testing methods. *Laryngoscope, 58*, 955-991.

Elliott, L. L. (1995). Verbal auditory closure and the speech perception in noise (SPIN) test. *Journal of Speech and Hearing Research, 38*, 1363-1376.

Festen, J., & Plomp, R. (1983). Relations between auditory functions in impaired hearing. *Journal of the Acoustical Society of America, 73*, 652-662.

Fletcher, H. (1929). *Speech and hearing.* Princeton, NJ: Van Nostrand.

Fletcher, H. (1950). A method of calculating hearing loss for speech from an audiogram. *Journal of the Acoustical Society of America, 22*, 1-5.

Fletcher, H. (1953). *Speech and hearing in communication.* Toronto: Van Nostrand.

Fournier, J. E. (1951). *L'audiométrie vocale.* Paris: Maloine.

French, N. R., & Steinberg, J. C. (1947.) Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America, 19*, 90-119.

Gatehouse, S. (1994). Components and determinants of hearing aid benefit. *Ear & Hearing, 15*, 30-49.

Gierut, J., & Pisoni, D. B. (1988). Speech perception. In N. J. Lass, L.V. Reynolds, J. L. Northern, & D. E., Yoder (Eds.), *Handbook of speech-language pathology and audiology* (pp. 253-276). Toronto: Decker.

Goetzinger, C. P. (1978). Speech discrimination testing. In J. Katz (Ed.), *Handbook of clinical audiology* (2nd ed., pp. 149-158). Baltimore: Williams & Wilkins.

Gordon-Salant, S. (1986). Effects of aging on response criteria in speech recognition tasks. *Journal of Speech and Hearing Research, 29*, 155-162.

Guilford, J. P. (1954). *Psychometric Methods* (2nd ed.). New York: McGraw-Hill.

Hébert, R., & Picard, M. (1979). *Test de discrimination de la parole.* Longueil: Prolingua.

Hirsh, I. J., Davis, H., Silverman, R. S., Reynolds, E. G., Eldert, E., & Benson, R.W. (1952). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders, 17*, 321-337.

Hopkinson, N. (1978). Speech reception threshold. In J. Katz (Ed.), *Handbook of clinical audiology* (2nd ed., pp. 141-148). Baltimore: Williams & Wilkins.

Hudghson, W., & Thompson, E. A. (1942). Correlation of hearing acuity for speech with discrete frequency audiograms. *Archives of Otolaringology, 36*, 526-540.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337-1351.

Keith, W. R. (1988). Central auditory tests. In N. J. Lass, L.V. Reynolds, J. L. Northern, & D. E. Yoder (Eds.), *Handbook of speech-language pathology and audiology* (pp. 1215-1236). Toronto: Decker.

Kent, R. D. (1992). Auditory processing of speech. In J., Katz, N., Stecker, & D. Henderson (Eds.), *Central auditory processing* (pp. 93-105). St-Louis, MO: Mosby.

Kruger, B., & Mazor, R. M. (1987). Speech audiometry in the USA. In M. Martin (Ed.), *Speech audiometry* (pp. 207-235). London: Whurr.

Lafon, J. C. (1958). *Le test phonétique.* Paris: Masson.

Lebel, C., & Picard, M. (1995). Développement et essai clinique du Test de seuil vocal en images (TSVI) pour enfants français québécois. *Journal of Speeech-Language Pathology and Audiology, 19*, 165-175.

Lebel, C., & Picard, M. (1997). Influence du mode de réponse sur le seuil de reconnaissance de la parole chez l'enfant français québécois. *Journal of Speeech-Language Pathology and Audiology, 21*, 17-27.

Lefebvre, S. (1991). *Dysfonctions auditives particulières chez des enfants avec histoire d'otites à répétition.* Unpublished Master's thesis, Université de Montréal, Montreal, Quebec.

Lemme, M. L., & Hedberg, N. L. (1988). Auditory linguistic processing. In N. J. Lass, L. V. Reynolds, J. L. Northern, & D. E., Yoder (Eds.), *Handbook of speech-language pathology and audiology* (pp. 277-308). Toronto: Decker.

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition. In M.A. Gernsbacker (Ed.), *Handbook of psycholinguistics* (pp. 265-301). New York: Academic Press.

Martin, F. N. (1978). Speech tests of hearing. In F. N. Martin (Ed.), *Pediatric audiology* (pp. 236-264). Englewood Cliffs: Prentice-Hall.

Martin, F. N. (1997). *Introduction to Audiology* (6th ed.). Englewoods Cliffs: Prentice-Hall.

Martin, F. N., & Morris, L. J. (1989). Current audiologic practices in the US. *Hearing Journal, 42*, 25-44.

Martin, F. N., & Sides, D. (1985). Survey of current audiometric practices. *ASHA, 27*, 29-36.

Nabelek, A. K., & Nabelek, I. V. (1994). Room acoustics and speech perception. In J. Katz (Ed.), *Handbook of clinical audiology* (4th ed., pp. 624-637). Baltimore: Williams & Wilkins.

Newby, H. A. (1979). *Audiology* (4th ed.). Englewood Cliffs: Prentice-Hall.

Nillson, M. J., Soli, S. D., & Sullivan J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95*, 1085-1099.

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *Journal of the Acoustical Society of America, 87*, 2705-2715.

Noble, W. G.(1978). *Assessment of impaired hearing*. New York: Academic Press.

Olsen, W. O. (1991). Special auditory tests: A historical perspective. In J. T. Jacobson & J. L. Northern (Eds.), *Diagnostic audiology*. Austin, TX: Pro-Ed.

Olsen, W. O., & Matkin, N. D. (1991). Speech audiometry. In W. F. Rintelmann (Ed.), *Hearing Assessment* (2nd ed., pp. 39-140). Boston: Allyn and Bacon.

Pavlovic, C. V. (1993). Problems in the prediction of speech recognition performance of normal-hearing and hearing-impaired individuals. In G. A. Studebaker & I. Hochberg (Eds.), *Acoustical factors affecting hearing aid performance* (2nd ed., pp. 221-234). Boston: Allyn and Bacon.

Penrod, J. P. (1994). Speech threshold and word recognition/discrimination testing. In J. Katz (Ed.), *Handbook of clinical audiology* (4th ed., pp. 147-164). Baltimore: Williams & Wilkins.

Picard, M. (1984). L'audiométrie vocale au Québec français. *Audiology, 23*, 337-365.

Picard, M., Banville, R., Barbarosie, T., & Manolache, M. (in press). Speech audiometry in noise-exposed workers: The SRT-PTA relationship revisited. *Audiology*.

Pichora-Fuller, M. K. (1996). Working memory and speechreading. In D. G. Stork & M. E. Hennecke (Eds.), *Speechreading by humans and machines*. *NATO ASI series* (pp. 257-274). Berlin: Springer Verlag.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97*, 593-608.

Pisoni, D.B., & Luce, P.A. (1987). Trading relations, acoustic cue integration, and context effects in speech perception. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 155-172). Dordrecht: Martines Nijhoff.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *Journal of the Acoustical Society of America, 63*, 533-549.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech and Hearing Research, 29*, 146-154.

Plomp, R. (1994). Noise, amplification, and compression. *Ear & Hearing, 15*, 2-12.

Plomp, R., & Mimpen, A.M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *Journal of the Acoustical Society of America*, 66, 1333-1342.

Ross, M., & Lerman, J. (1970). A picture identification test for hearing-impaired children. *Journal of Speech and Hearing Research, 13*, 44-53.

Rupp, R. R. (1980). Classical approaches to the determination of the spondee threshold. In R. R. Rupp & K.G., Stockdell (Eds.), *Speech protocols in audiology* (pp. 67-97). New York: Grune & Stratton.

Schill, H. A. (1985). Thresholds for speech. In J. Katz (Ed.), *Handbook of clinical audiology* (3rd ed., pp. 224-234)). Baltimore: Williams & Wilkins.

Silman, S., & Silverman, C. A. (1991). *Auditory diagnostic: Principles and applications*. Toronto: Academic Press.

Simpson, G. B. (1994). Context and the processing of ambiguous words. In M. A. Gernsbacker (Ed.), *Handbook of psycholinguistics* (pp. 359-374). New York: Academic Press.

Stockdell, K. G. (1980). Classical approaches to measuring discrimination efficiency via word lists. In R. R. Rupp & K.G. Stockdell (Eds.), *Speech protocols in audiology* (pp. 99-118). New York: Grune & Stratton.

Strange, W. (1986). Speech input and the development of speech perception. In J.F. Kavanagh (Ed.), *Otitis media and child development* (pp. 12-26). Parkton: York Press.

Studebaker, G.A., & Sherbecoe, R.L. (1993). Frequency-importance functions for speech recognition. In G. A. Studebaker & I. Hochberg (Eds.), *Acoustical factors affecting hearing aid performance* (2nd ed., pp. 185-204). Boston: Allyn and Bacon.

Takata, Y., & Nabelek, A.K. (1990). English recognition in noise and in reverberation by Japanese and American listeners. *Journal of the Acoustical Society of America*, 88, 663-666.

Thornton, A. R., & Raffin, M. J. (1978). Speech-discrimination scores modeled as a binomial variable. *Journal of Speech and Hearing Research, 21*, 507-518.

Tillman, T. W., & Olsen, W. O. (1973). Speech audiometry. In J. J. Jerger (Ed.), *Modern development in audiology* (2nd ed., pp. 37-74). New York: Academic Press.

## *Appendix C*

### Word lists from Lebel and Picard (1995) for the measurement of SRT in children.

| List A | List B | List C | List D | List E |
|---|---|---|---|---|
| 1. Lapin | bateau | cadeau | oiseau | poisson |
| 2. Oiseau | ciseaux | bateau | soleil | ciseaux |
| 3. Ciseaux | poisson | cochon | cochon | fourchette |
| 4. Poisson | fourchette | oiseau | ciseaux | cheval |
| 5. Cuillère | chandelle | chandelle | poisson | camion |
| 6. Camion | oiseau | camion | cadeau | cuillère |
| 7. Chandelle | soleil | poisson | chandelle | oiseau |
| 8. Avion | camion | mitaines | cuillère | sapin |
| 9. Carotte | cheval | lapin | lapin | cochon |
| 10. Cheval | cochon | fourchette | carotte | avion |
| 11. Fourchette | ballon | ciseaux | ballon | ballon |
| 12. Ballon | sapin | soleil | fourchette | mitaines |
| 13. Cochon | cuillère | cuillère | avion | soleil |
| 14. Sapin | avion | ballon | sapin | chandelle |
| 15. Cadeau | lapin | sapin | cheval | carotte |
| 16. Bateau | mitaines | carotte | mitaines | lapin |
| 17. Mitaines | cadeau | cheval | camion | cadeau |
| 18. Soleil | carotte | avion | bateau | bateau |

## *Appendix D*

### Word lists from Hébert-Picard (1979) for the measurement of PWRS in children.

| List 1 | List 2 | List 3 | List 4 |
|---|---|---|---|
| 1. Gant | camp | banc | dent |
| 2. Roue | loup | sou | joue |
| 3. Fer | verre | mère | terre |
| 4. Loup | zoo | joue | sou |
| 5. Main | bain | faim | pain |
| 6. Saigne | sept | cenne* | sel |
| 7. Sou | joue | roue | chou |
| 8. Dire | tire | cire | lire |
| 9. Voir | boire | poire | noir |
| 10. Mousse | pouce | tousse | bourse |
| 11. Fondre | montre | pondre | tondre |
| 12. Jappe | cape | nappe | tape |
| 13. Beurre | coeur | peur | soeur |
| 14. Tache | vache | hache | cache |
| 15. Jeu | feu | queue | boeufs |
| 16. Boeufs | banc | bas | bain |
| 17. Balle | boule | bol | belle |
| 18. Chaud | chat | champ | chou |
| 19. Rue | rond | roue | rit |
| 20. Fée | feu | faim | fond |

*Old French meaning 'penny'.

## *Appendix E*

### Lists of high(H) and low(L) predictability sentences for the measurement of sentence-level context effects in children (Lefebvre, 1991).

| List | | Predictability |
|---|---|---|
| **3A** | 1. Anne a une coupure sur la joue | H |
| | 2. L'autobus a perdu une roue | H |
| | 3. Mon papa croque le banc | L |
| | 4. Le chien s'est fait couper la queue | H |
| | 5. Jean écoute le chant des cennes | L |
| | 6. Après avoir couru, j'ai faim | H |
| | 7. La piscine est pleine de poires | L |
| | 8. L'hiver, on doit mettre nos bas | H |
| | 9. Le chat est plein de roues | L |
| | 10. L'autobus roule souvent sa mère | L |
| | 11. La chandelle est faite de cire | H |
| | 12. La neige fond quand elle a peur | L |
| | 13. Le sapin est rempli de faim | L |
| | 14. Luc coupe l'arbre avec sa hache | H |
| | 15. L'auto fonctionne avec une nappe | L |
| | 16. Le blé pousse dans le champ | H |
| | 17. La poule vient juste de pondre | H |
| | 18. En famille, on mange de la tousse | L |
| | 19. Dans ma poche, il y a un sou | H |
| | 20. Les enfants courent dans le bol | L |
| **3B** | 1. Le matin, Julie a très faim | H |
| | 2. Le lion rugit de cire | L |
| | 3. Le pain est fait avec du champ | L |
| | 4. Le bébé s'ennuie de sa mère | H |
| | 5. Les érables poussent dans la roue | L |
| | 6. Les céréales sont dans un bol | H |
| | 7. Le pingouin s'envole dans la hache | L |
| | 8. Ma bicyclette a deux roues | H |
| | 9. Le professeur enseigne aux bas | L |
| | 10. Alain veut des bonbons pour cinq cennes | H |
| | 11. Jean a mis son sac sur le banc | H |
| | 12. L'oiseau boit des sous | L |
| | 13. La table est sur ma joue | L |
| | 14. Le chat se sauve, car il a peur | H |
| | 15. Ma soeur a le rhume et tousse | H |
| | 16. Le panier est rempli de poires | H |
| | 17. Le cheval regarde la pondre | L |
| | 18. Pierre met de l'essence dans la queue | L |
| | 19. Sur la table, il y a une nappe | H |
| | 20. L'été, je nage dans la faim | L |

Tremblay, C., Picard, M., Barbarosie, T., & Banville, R. (1991). Etude clinique des décalages de seuil vocal dans le bruit. *Audiology*, *30*, 214-240.

Turner, R. G. (1988). Technique to determine test protocol performance. *Ear and Hearing*, 9, 177-189.

Turner, R. G. (1991). Making clinical decision. In W.F. Rintelmann (Ed.), *Hearing Assessment* (2nd ed., pp. 679-738). Boston: Allyn and Bacon.

Turner, R. G., Shepard, N. T., & Frazer, G. J. (1984). Clinical performance of audiological and related diagnostic tests. *Ear and Hearing*, 5, 187-194.

Ventry, I. M. (1976). Pure tone-spondee threshold relationships in functional hearing loss: a hypothesis. *Journal of Speech and Hearing Disorders*, *41*, 16-22.

Wilson, R. H., & Margolis, R. H. (1983). Measurements of auditory thresholds for speech stimuli. In D.F. Konkle & W.F., Rintelmann (Eds.), *Principles of speech audiometry* (pp. 79-126). Baltimore: University Park Press.

## *Appendix A*

**Word lists A, B, C, and D as suggested by Picard (1984) for the measurement of SRT in adults (11 highly recognizable bisyllables).**

| List A | List B | List C | List D |
|---|---|---|---|
| 1. Programme | orange | modèle | parole |
| 2. Couloir | modèle | carotte | horloge |
| 3. Fromage | horloge | docteur | programme |
| 4. Docteur | programme | couloir | départ |
| 5. Départ | carotte | parole | docteur |
| 6. Horloge | docteur | départ | voiture |
| 7. Carotte | parole | orange | modèle |
| 8. Parole | voiture | programme | couloir |
| 9. Orange | départ | fromage | carotte |
| 10. Voiture | couloir | horloge | fromage |
| 11. Modèle | fromage | voiture | orange |
| 12. Fromage | programme | parole | modèle |
| 13. Horloge | carotte | docteur | parole |
| 14. Modèle | fromage | orange | couloir |
| 15. Parole | modèle | programme | voiture |
| 16. Départ | docteur | modèle | fromage |
| 17. Voiture | orange | départ | docteur |
| 18. Couloir | voiture | carotte | horloge |
| 19. Carotte | parole | voiture | programme |
| 20. Docteur | horloge | fromage | orange |
| 21. Orange | départ | couloir | carotte |
| 22. Programme | couloir | horloge | départ |

## *Appendix B*

**Word lists A, B, E, and F from Benfante et al. (1966) for the measurement of PWRS in adults.**

| List A | List B | List E | List F |
|---|---|---|---|
| 1. bar | base | berne | salle |
| 2. gel | chic | quel | taire |
| 3. celle | mince | caisse | cher |
| 4. crème | range | seize | sèche |
| 5. guêpe | veuf | phare | pâle |
| 6. cor | bile | harpe | gaz |
| 7. laide | ligue | bague | page |
| 8. juche | chose | bol | casse |
| 9. pose | dôme | dock | bonne |
| 10. taupe | sauf | phoque | molle |
| 11. pour | ruche | pomme | roule |
| 12. coq | mule | rousse | loupe |
| 13. vole | doute | touche | soute |
| 14. orgue | coûte | bouge | louche |
| 15. moule | bouche | boule | mur |
| 16. joute | douze | nuque | bulle |
| 17. fougue | toque | saule | puce |
| 18. fume | loge | sud | chaume |
| 19. neige | face | cause | sauve |
| 20. chaude | vague | geôle | tige |
| 21. cil | gare | ville | hymne |
| 22. fiche | mèche | guise | rose |
| 23. soeur | cerf | beurre | banque |
| 24. rance | herbe | chante | change |
| 25. chatte | homme | jambe | gomme |
| 26. mise | fuite | folle | donne |
| 27. cogne | poil | Pâques | signe |
| 28. prince | transe | vigne | pierre |
| 29. pièce | blonde | nuire | tuile |
| 30. plus | grippe | ciel | coiffe |
| 31. râcle | griffe | moite | contre |
| 32. phrase | cirque | cintre | fleur |
| 33. quatre | givre | fondre | vitre |
| 34. charme | prune | risque | disque |
| 35. sable | brute | livre | fibre |
| 36. solde | couple | tigre | frise |
| 37. forte | course | chiffre | luxe |
| 38. borne | snob | plume | foudre |
| 39. bourse | flore | sucre | groupe |
| 40. fourbe | morte | moudre | tourbe |
| 41. trouve | sport | fourche | gorge |
| 42. cruche | parle | troupe | stock |
| 43. bigle | charge | porche | cloche |
| 44. style | casque | grotte | drogue |
| 45. crime | boucle | morse | parc |
| 46. brique | lettre | large | barbe |
| 47. trombe | clair | classe | vaincre |
| 48. grande | plaire | blaque | lèvre |
| 49. voile | nièce | perche | ferme |
| 50. ruine | rogne | maître | blesse |