This paper was published in: *Proceedings of the 1997 IEEE Workshop on Speech Coding for Telecommunications*, Pocono Manor, PA, September 1997.

LISTENER RATINGS OF SPEECH PASSBANDS

Stephen Voran

Institute for Telecommunication Sciences, U.S. Department of Commerce, NTIA/ITS.N3
325 Broadway, Boulder, Colorado 80303, sv@bldrdoc.gov

ABSTRACT

We describe a listening experiment that measures the perceived speech quality of 19 speech passbands using 8 talkers and 28 listeners. Results are referenced to the traditional wide-band and narrow-band telephony passbands. Our findings may help those who wish to select new speech coding passbands that maximize perceived speech quality under bit-rate constraints. We identify several passbands that show particular promise.

1. BACKGROUND

One of the most basic design decisions in speech coding is the selection of a passband to code. Compatibility with the analog phone network dictated the use of the 300-3400 Hz passband in early PCM systems, as specified in CCITT Recommendation G.712. This passband has become a de facto definition in "narrow-band telephony"(NB). Similarly, the 50-7000 Hz passband specified in CCITT Recommendation G.722 has become a de facto definition in "wide-band telephony"(WB). Compatibility with these passbands is often important, but there are also situations where compatibility may not be necessary, desirable, or even possible. When all other parameters are held constant, coding a wider passband generally produces superior speech quality, but at the cost of increased bit rate. We have performed a listening experiment that measures the perceived speech quality of 19 speech passbands, including the NB and WB cases. Our results may guide those who wish to select new coding passbands that maximize speech quality under bit-rate constraints.

2. EXPERIMENT DESIGN

We wish to determine the perceived speech quality of different coding passbands under a constant bit-rate constraint. We claim that for a fixed coding technique, a constant bit-rate constraint can be approximated with a constant coding bandwidth constraint, when coding bandwidth is measured in critical bands or in Bark (B). A survey of frequency-domain bit-allocation techniques in speech and audio coders shows that this is a reasonable claim. Thus we considered families of passbands that have constant bandwidths of 11.7, 14.1, and 16.0 B, as defined in Table 1. These bandwidths span the range of bandwidths between the NB (11.7 B) and WB (18.4 B) cases. For each bandwidth, four or five passbands are defined, starting at 50 Hz, and shifting up in 0.8 B increments. We use the Bark-to-Hz transformation $h=600\cdot\sinh(b/6)$[1]. Passband 4 is the NB case and Passband 15 is the WB case.

In uncoded digital transmission, a constant bit-rate constraint is equivalent to a constant speech bandwidth constraint, when speech bandwidth is measured in Hertz. Similarly, in analog transmission, a constant channel bandwidth constraint is equivalent to a constant speech bandwidth (Hertz) constraint. Thus we defined Passbands 16-19 to allow the study of passband options under the 3100 Hz bandwidth constraint of the NB case.

We designed a paired-comparison listening experiment to determine the perceived speech quality of these 19 passbands relative to the NB case. Single-sentence recordings from four female and four male English language talkers were used. These digital recordings use a sample rate of 16 kHz, have 7.5 kHz bandwidth, and SNR's between 44 and 53 dB. The lowest observed pitches range from 160 to 180 Hz for the female talkers, and from 90 to 100 Hz for the male talkers. The total duration of these recordings is 25 seconds. Each of the eight recordings was processed by 19 software bandpass filters, with -3 dB points as specified in Table 1. We used order 300 FIR filters, with passband ripple below ±.02 dB, transition bands that are 100 to 120 Hz wide and typical stopband attenuation of 55 dB. The 152 resulting recordings were then normalized to a common level using a software approximation of an A-weighted sound level meter.

For each of the 8 talkers, we formed 19 pairs of recordings, using Passband 4 (NB) as a reference in each pair. Pair ordering was randomized under the constraint that the reference appear first in 50% of the pairs for each talker. The 152 pairs were presented in a different random order to each of 28 listeners. Thus, each passband was evaluated 224 times (8 talkers x 28 listeners). The 14 female and 14 male listeners range in age from 25 to 65 years. The average female or male age is 44 years. Pure-tone air-conduction hearing thresholds were measured for all listeners at eight frequencies between 125 Hz and 8 kHz. The measured thresholds are within the expected limits for listeners in this age range.

The experiment was conducted in a sound booth with ambient noise below noise criterion 30. Listeners wore studio monitor headphones, connected to a digital playback system with 70 dB dynamic range, and nearly flat frequency response that rolls off to -1.3 dB at 50 Hz and -1.0 dB at 7 kHz. After hearing each pair of recordings, each listener used an electronic screen and pen system to select one of the seven options in the middle of this sentence: "The second version sounds much better than, better than, slightly better than, the same as, slightly worse than, worse than, much worse than, the first version." For analysis purposes, these seven options were later assigned the scores 3, 2,…,-3 respectively, when the reference was the first element of the presentation pair. When the reference was the second element of the presentation pair, these scores were negated.

These passbands might also be compared using intelligibility, listening effort, or task performance. The articulation index predicts that for this noise-free case, word intelligibility in sentence context will increase from 99.3% to 99.9% when NB is replaced with WB[2,3]. Thus it is highly unlikely that word intelligibility in sentence context would differentiate passbands in this type of experiment, unless a family of relevant acoustic noise environments were defined and implemented. We expect that the same would be true for listening effort and task performance experiments.

3. RESULTS AND OBSERVATIONS

Table 1 gives the mean score (across all talkers and listeners) and the half-width of its 95% confidence interval (CI) for each passband. We make an implicit scale linearity assumption when we make these calculations. Results for the 11.7 B case are also shown in Figure 1. Of the five 11.7 B wide passbands, Passband 3 is preferred. Because Passband 4 (NB) is also the reference, its mean score is near zero and has a relatively narrow confidence interval. Passbands 1, 2, and 5 have negative means, indicating that on average, they sound worse than the NB reference. Similar graphs can be drawn for the 14.1 B, 16.0 B, and 3100 Hz cases. The highest scores for these cases are associated with Passbands 7, 11, and 17 respectively, but some confidence intervals overlap in these cases. In all three cases, passbands that extend below 300 Hz are preferred to those that do not. Up to a point, the low frequencies gained seem to outweigh the high frequencies lost. The results of the 3100 Hz case confirm the value of a technique used by broadcasters who gather program material using analog phone lines. They employ "frequency extenders" that shift frequencies below 300 Hz up into the phone line passband, and back down again at the receiving end.

An examination of Passbands 4, 6, 14, and 15, provides insight into the perceived speech quality improvement offered by WB over NB. 61% of the score increase can be had by extending the lower limit of the passband alone (Passband 6). Extending the upper limit alone (Passband 14) has no apparent advantage, but that extension clearly complements the extension of the lower limit (Passband 15).

Figure 2 shows the highest scoring passbands for each of the constant Bark bandwidths in this experiment. This graph of perceived speech quality versus bandwidth (or bit rate) reveals several interesting results. Passband 3 demonstrates that 20% of the score increase associated with WB over NB can be had without increasing the 11.7 B bandwidth of NB at all. 77% of the WB improvement can be achieved by a properly positioned 14.1 B wide passband (Passband 7), and 94% of the WB improvement can be obtained by correctly locating a 16.0 B passband (Passband 11).

The results reported here may provide guidance to those who wish to select new coding passbands that maximize perceived speech quality under fixed bit-rate constraints. In particular, we have located passbands that may offer near WB quality at reduced bit rates, and have identified promising alternative locations for the 11.7 B and 3100 Hz bandwidths of NB. These are fundamental results on flat passbands in the absence of coding distortions. Further experiments may identify interactions among passband limits, passband shape, coding distortions, and perceived speech quality.

Pass-band	Lower Limit	Upper Limit	Band-width	Mean Score	95% CI (±)
1	50 Hz	2262 Hz	11.7 Bark	-1.42	0.19
2	131	2594	"	-0.29	0.21
3	213	2971	"	0.29	0.16
4 (NB)	300	3400	"	-0.01	0.06
5	392	3889	"	-1.00	0.15
6	50	3400	14.1 Bark	0.86	0.16
7	131	3889	"	1.10	0.14
8	213	4447	"	0.94	0.15
9	300	5083	"	0.16	0.14
10	392	5809	"	-1.31	0.15
11	50	4691	16.0 Bark	1.34	0.15
12	131	5362	"	1.24	0.16
13	213	6127	"	1.12	0.15
14	300	7000	"	-0.06	0.18
15(WB)	50	7000	18.4 Bark	1.42	0.16
16	50	3150	3.1 kHz	0.58	0.17
17	131	3231	"	0.63	0.16
18	213	3313	"	0.59	0.14
19	392	3492	"	-0.86	0.12

Table 1. Passband Definitions and Scores

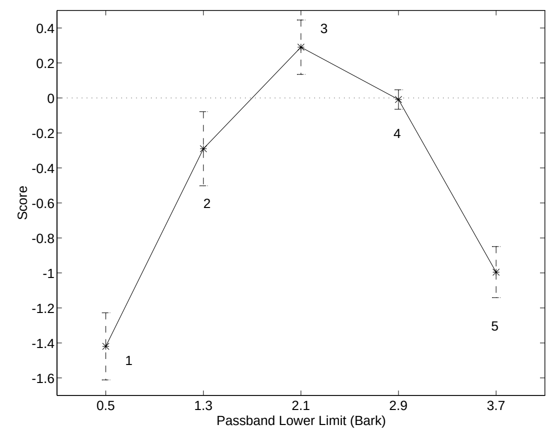

Figure 1. Means and 95% CI's for 11.7 B Passbands

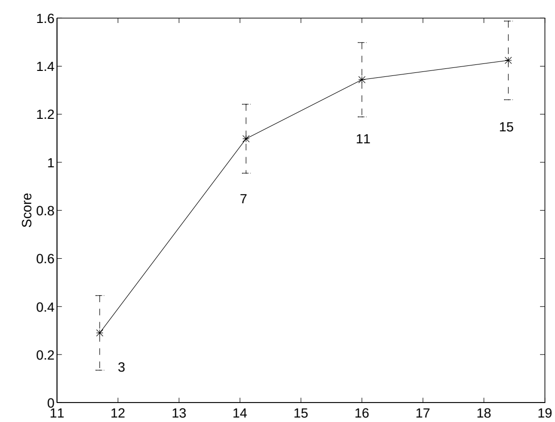

Figure 2. Highest Rated Passbands

REFERENCES

[1] A.J. Fourcin, et al., "Speech processing by man and machine," in *Recognition of Complex Acoustic Signals*, T.H. Bullock, Ed., Report of Dahlem Workshop, Berlin, 1977.

[2] J.B. Allen, "How do humans process and recognize speech?" IEEE Trans. Speech and Audio Proc., Vol. 2, Oct. 1994.

[3] A. Boothroyd & S. Nittrouer, "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am.,Vol. 84, Jul. 1988.