J Am Acad Audiol 18:688–699 (2007)

Speech and Music Quality Ratings for Linear and Nonlinear Hearing Aid Circuitry

Evelyn Davies-Venn*
Pamela Souza*
David Fabry†

Abstract

This study evaluated quality ratings for speech and music stimuli processed using peak clipping (PC), compression limiting (CL), and wide-dynamic range compression (WDRC) hearing aid circuitry. Eighteen listeners with mild-to-moderate hearing loss were binaurally fitted with behind-the-ear (BTE) hearing aids and instructed to rate the quality of speech under various conditions in quiet and noise and two genres of music. Results for speech revealed a slight preference for WDRC at 80 dB SPL, and equivalent ratings for the three circuits under all other listening conditions. Music ratings revealed a marginally significant preference for WDRC and a preference for classical over popular music. For music, judgments on pleasantness were the most influential on overall circuit preference.

Key Words: Hearing aids, peak clipping, compression, speech recognition, HINT, quality

Abbreviations: BTE = behind-the-ear; CL = compression limiting; PC = peak clipping; PTA = pure-tone average; SNR = signal to noise ratio; WDRC = wide -dynamic range compression.

Sumario

Este estudio evalúa estimaciones de calidad para estímulos de lenguaje y música procesados usando corte de picos (PC), limitación de la compresión (CL) y compresión de rango dinámico amplio (WDRC) en los circuitos del auxiliar auditivo. Dieciocho sujetos con hipoacusia leve a moderada fueron adaptados binauralmente con auxiliares auditivos retroauriculares (BTE) e instruidos para juzgar la calidad del lenguaje bajo varias condiciones de silencio y ruido, y la de dos géneros de música. Los resultados para el lenguaje revelaron una ligera preferencia para el WDRC a 80 dB SPL, y estimaciones equivalentes para los tres circuitos en las otras condiciones de escucha. Las estimaciones para la música revelaron una preferencia marginalmente significativa para el WDRC y una preferencia para la música clásica por encima de la popular. Para la música, los juicios de agradabilidad fueron los que más influyeron en la preferencia global de los circuitos.

Palabras Clave: Auxiliares auditivos, corte de picos, compresión, reconocimiento del lenguaje, HINT, calidad

Abreviaturas: BTE = retroauricular; CL = limitación de la compresión; PC = corte de picos; PTA = promedio tonal puro; SNR = tasa señal/ruido; WDRC = compresión de rango dinámico amplio

*University of Washington, Seattle; †Phonak Inc, Warrenville, IL

Dr. Pamela Souza, Department of Speech & Hearing Sciences, University of Washington, 1417 NE 42nd Street, Seattle, WA 98105

Although current hearing instruments offer a range of features for the hearing impaired user, one of the most basic decisions made by clinicians is whether to provide a linear or nonlinear amplification strategy. Linear amplification applies a constant gain for low- to mid-level input signals. For high-level input signals, the gain is reduced by either compression limiting (CL) or peak clipping (PC). In digital hearing aids, nonlinear amplification usually implies multichannel wide-dynamic range compression. This amplification strategy is intended to improve speech intelligibility by using fast, automatic gain adjustments to place more speech cues within a listener's audible range and also provide progressively less gain at higher input levels to aid listening comfort for loud sounds.

Although there are many studies comparing WDRC versus PC amplification, or WDRC versus CL amplification, fewer investigators made direct comparisons among WDRC, CL, and PC processing in the same subject cohort (e.g., Larson et al, 2000; Lunner et al 1997, 1998). The majority of these comparisons prioritized speech recognition, with quality as a secondary consideration. As a whole, previous work indicated no significant difference for speech recognition performance in quiet and noise as well as speech quality on all dimensions, except loudness, where WDRC was usually superior (e.g., Jenstad et al, 2000; Larson et al, 2000; Lunner et al, 1998; Souza et al, 2002).

Most previous work also focused exclusively on speech. However, hearing aid wearers reported that other signals are also important. In a previous survey, three quarters of hearing aid wearers reported using their aid to listen to music (Feldmann and Kumpf, 1988). Although most hearing aid users were satisfied with the response of their aid for music listening (Kochkin, 2000a), it was less clear whether some types of signal processing are superior to others for music listening. We do know that multichannel compression of the sound envelope reduced music quality (van Buuren et al, 1999), which might mean that a linear response is preferred over multichannel WDRC. On the other hand, Chasin (2003) noted that the spectrum of music is much more variable compared to that of speech. Unlike the resonators of the speech production system, hard walled instruments used for musical productions usually result in crest factors ranging from 18 to 20 dB, compared to approximately 12 dB for speech. That suggests that WDRC could improve audibility of low-level musical notes that would otherwise be inaudible to the listener, and in that respect might be preferred for musical selections with wide intensity variations.

We are aware of only a few studies that included music in their quality judgment task (Franks, 1982; Lunner et al, 1997; Punch, 1978). Among these studies, only Lunner and colleagues (1997) made a direct comparison of all three amplification strategies. However, their study cumulated ratings across various listening environments, rather than analyzing amplification strategy preference for music alone. Previous data indicated that amplification parameters which optimize speech quality did not necessarily predict optimal music quality (van Buuren et al, 1996). Thus, there is a need for studies evaluating listener preference for amplification strategies for music (Chasin and Russo, 2004), in the context of quality ratings and objective performance for speech in quiet and noise.

Finally, although previous studies evaluated the relationship between distinct perceptual dimensions of speech quality and overall speech intelligibility ratings (Gabrielsson et al, 1988; Preminger and Van Tasell, 1995a, 1995b), little is known about the influence of specific aspects of music quality on overall impression ratings for music. For example, we do not know whether the dimension that best predicts overall preference differs for speech compared to music listening. If speech and music share the same predictor for overall preference, then less clinical time may be spent measuring and adjusting a hearing aid to accommodate other quality dimensions that have minimal influence on a hearing aid wearer's overall preference. Rather, this knowledge may facilitate more efficient use of time by only focusing on the "best predictor" dimensions during hearing aid fitting appointments. Accordingly, the present study aimed to address these issues among a cohort of listeners with mild-to-moderate sensorineural hearing loss.

METHODS

Subjects

Eighteen adults with bilateral sensorineural hearing loss participated in this study. These adults were between 28 and 86 years old. Participants' mean age was 69.8 years (SD = 13.58). Except for a few subjects who reported some piano or voice lessons, all participants only listened to music for pleasure. None of the participants had professional music training. Half of the study participants had at least six months of full time hearing aid use (3 to 25 years of hearing aid experience), and half had no experience listening through a hearing aid. Mean audiometric thresholds (in dB HL) are shown in Table 1 and subject demographics in Table 2.

English was the primary language of all participants. Selection was based on satisfying specific inclusion criteria: a sensorineural hearing loss with thresholds between 25-85 dB HL at octave frequencies from .25 to 8 kHz; binaural word recognition score 80% or better at 30 dB SL re: SRT, using a recorded version of the Northwestern University (NU-6) monosyllables (Wilson et al, 1990). The exclusion criteria included: presence of an air-bone gap that exceeded 10 dB at any frequency, asymmetry between ears greater than 20 dB at any frequency from 1 kHz to 8 kHz, word recognition scores less than 80%, the presence of middle ear involvement as indicated by any shape other than the normal Type A 226 Hz tympanogram (Jerger, 1970), or retrocochlear pathology based on self-reported medical history. During the

Table 1. Mean Hearing Thresholds in Decibel Hearing Level (dB HL)

Frequency (Hz)	Right ear	SD	Left ear	SD
250	25.3	5.3	23.9	7.4
500	31.1	7.2	29.7	8.5
1000	39.4	8.2	37.8	10.6
2000	48.3	8.6	47.8	9.3
3000	52.8	11.4	53.6	9.8
4000	55.0	12.0	56.9	11.0
6000	61.2	11.1	63.2	11.9
8000	64.7	14.1	64.2	14.6
PTA	39.6	6.6	38.4	8.5

Table 2. Subject Demographics

Subject	Age (yrs)	Aid type	Aid style	Yrs of HA use	NU-6 score*
1	73	CL	ITE	25	98
2	71	WDRC	ITE	9	92
3	52	WDRC	BTE	3	94
4	28	CL	CIC	22	98
5	79	CL	CIC	3	94
6	68	CL	CIC	2	90
7	59	WDRC	BTE	11	92
8	64	WDRC	ITC	12	98
9	77	CL	ITC	5	90
10	74	None			94
11	75	None			98
12	81	None			88
13	62	None			96
14	73	None			94
15	86	None			89
16	84	None			81
17	75	None			98
18	76	None			90

*Average of monaural-right and monaural-left NU6 scores.

first visit, all qualified participants were also administered loudness discomfort testing for warble tones at .5 and 3 kHz using instructions recommended by Hawkins et al (1987). Specifically, the instructions and stimuli were those suggested by Hawkins et al., but participants were tested using insert earphones (ER3A). Individual earmold impressions were made following the audiometric testing.

The second visit included hearing aid fitting, followed by aided speech testing as described below. All participants were binaurally fit with Phonak Valeo 211 AZ digital behind-the-ear (BTE) hearing aids using custom-made, lucite skeleton earmolds with standard #13 tubing and an appropriate select-a-vent (SAV) size ranging from unvented to 2 mm. The filter bank in Valeo includes 15 independent channels, each 500 Hz wide, with filter slopes in excess of 50 dB/octave. The average level of input signals is calculated by a twin average time constants: a slow average detector and a fast average detector. Both are always active, but only one is in control of the gain at any moment in time, providing an adaptive compression attack/ release time that may vary independently across compression channels, with measured 2-cc coupler attack and release times of 40 ms and 120 ms, respectively. Each of the 15 compression channels has its own expansion, threshold kneepoint (TK), compression ratio (CR), and maximum power output (MPO). Digital noise cancellation, feedback management, and feedback phase inversion are independently active in each channel. Peak clipping, when used, is achieved via instantaneous (0 ms) attack/ release times, assuring that the defined output limit (MPO) is never exceeded (e.g., "hard limiting").

For every participant, each program memory of the experimental hearing instrument was programmed with a different signal processing strategy. One program was set as peak clipping (PC), one as compression limiting (CL), and one as wide-dynamic range compression (WDRC). All programs were set to omni-directional mode. The volume wheel, program button, noise suppression, and feedback suppression were deactivated in all three program memories. Expansion was available in all three programs and could not be deactivated. Expansion ratios were 0.65 in the PC and CL program and ranged from 0.5 to 0.7 (depending on channel) in the WDRC program.

Real-ear analysis was used to verify fittings for each of the three processing strategies, using standard clinical procedures (Mueller, 1992). National Acoustic Laboratories–Revised (NAL-R; Byrne and Dillon, 1986) real-ear insertion gain targets were generated for each subject. Although one memory used WDRC processing, multi-level targets were not generated due to limitations of the real-ear system used (Fonix 6500). Instead, the frequency-gain response for each program was adjusted until the mid-level (65 dB SPL input) response closely approximated the NAL-R target. Thus, the same mid-level input prescriptive target was used for all three signal processing strategies to control for audibility. This allowed direct inference to be made about difference in subject's performance for the experimental processing strategies, and reduced the possibility that difference in prescribed target, rather than signal processing, influenced either participant's objective performance or their subjective ratings. Mean real-ear gain was within 5 dB of NAL-R prescribed target gain for most subjects and frequencies (Figure 1). The remainder of the amplification parameters including expansion ratio, compression threshold, compression ratio, and MPO were individually set to the manufacturer's recommended settings for that audiogram. The compression threshold for the WDRC program was 39 dB SPL for each subject. Compression ratios for the WDRC program are shown in Table 3.

Finally, real ear saturation response was measured in each program using a 90 dB SPL pure-tone sweep and compared to individually measured loudness discomfort levels. Measured loudness discomfort levels were obtained in dB HL and later converted to dB SPL via the real ear system (Fonix 6500). If the saturation response exceeded individual loudness discomfort levels, the maximum output was reduced accordingly.

Speech Quality

The Speech Intelligibility Rating Test (SIR) developed by Cox and McDaniel (1984, 1989) was used for evaluating subjective speech quality ratings. Speech was presented at 50, 65, and 80 dB SPL in quiet, representing soft, conversational, and loud speech, respectively. Two noise conditions, one with

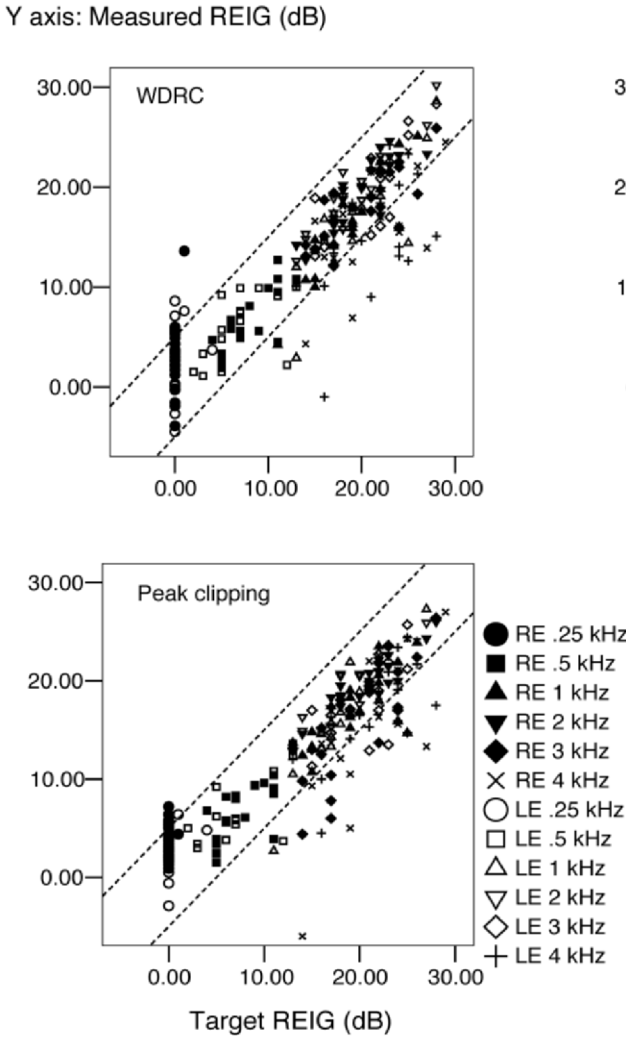

Figure 1. Measured real-ear insertion gain (REIG) as a function of target REIG. Results are shown for the right and left ears of each subject, and for each test frequency. The dashed lines in each panel indicate a deviation of +/- 5 dB relative to target.

speech at 65 dB SPL and the SIR multitalker babble at 55 dB SPL (+10 SNR), and the other with speech at 65 dB SPL and babble at 59 dB SPL (+6 dB SNR) were also included, for a total of five speech conditions. All SIR sentences are designed to be representative of everyday/continuous discourse. Although the developers suggest cueing subjects about each passage's topic prior to sentence presentation, subjects in this study were not cued on the topic. Listeners were also not required to repeat the sentences; rather, they were instructed to listen to the first five sentences and simply evaluate the sentences on four specified dimensions of speech quality. After the fifth sentence, subjects were cued to commence their quality ratings. Thus, subjects made these ratings while listening to the remaining five sentences in each passage. This method was used to minimize taxing subject's memory or their ability to recall the perceived quality of all 10 sentences after

Table 3. Compression Ratios at .5, 1, and 2 kHz across the 18 Subjects

Ear	.5 kHz	1 kHz	2 kHz
Right (mean)	1.2	1.56	1.59
Right (range)	1.1 – 1.3	1.4 – 1.7	1.3 – 1.9
Left (mean)	1.19	1.52	1.56
Left (range)	1.1 – 1.4	1.4 – 1.7	1.3 – 1.9

Note: Values were obtained from the fitting software and were not measured empirically.

presentation.

Subjects rated four aspects of speech quality, using a 1-10 scale: overall impression, pleasantness, intelligibility, and loudness. Definitions (Appendix A) were adapted from those used by Gabrielsson et al. (1998). The highest rating (10) represented the optimum rating for all of the scales except loudness, which ranged from "not loud at all" (0) to "very loud" (10) with 5 as the optimum rating. Ratings were made for each of the three amplification conditions and five different signals. The order of signal processing strategy was randomized, and the order of the five test signals was randomized within each signal processing condition. A different randomly selected passage was used for each condition.

Music Quality

One-minute segments of an instrumental piece by Mozart, Serenade no. 13 in G Major (Menuetto: Allegro), and a vocal piece by Virginia Rodriguez, "Adeus Batucada," were used to assess quality ratings of music presented at 65 dB SPL in quiet. Music ratings immediately followed the speech rating segment. Again, the order of signal processing strategy was randomized. Within each signal processing condition each subject first listened to a one-minute segment of an instrumental classical piece, followed by a one-minute segment of the vocal selection. Because intelligibility was not a specified dimension, the vocal piece was sung in Portuguese to reduce the likelihood of listeners rating this music for comprehension rather than quality. Similar to the speech rating segment, subjects were instructed to listen for 30 seconds and then rate the music while listening to the remaining 30 seconds of the music, using a rating scale of 1 to 10. The dimensions used for music rating were loudness, sharpness, fullness, pleasantness, and overall impression (Appendix B).

Speech Recognition

Although our primary interest was sound quality, it would be undesirable to compromise speech intelligibility. Accordingly, objective speech recognition performance in noise was evaluated using the Hearing in Noise Test (HINT) (Nilsson et al, 1994). This test uses an adaptive procedure to determine threshold for speech presented in speech-spectrum noise. For this experiment, speakers (JBL LSR25P) were positioned at 0 and 180 degrees azimuth, each 1 m from the listener's position.[1] Speech was presented from the front speaker and noise from the back speaker. This configuration has been shown to provide a realistic estimate of real-life speech in noise as long as omnidirectional microphones are used (Compton-Conley et al, 2004). In all other respects, the instructions, administration, and scoring method were as recommended by the developers of the HINT test (Nilsson et al, 1994). Presentation order of the three signal processing strategies was randomized for each subject.

RESULTS

Speech Quality

Mean ratings for the four quality dimensions are shown in Table 4. Rating scores for each dimension were analyzed using a two-way repeated measures ANOVA.[2] For speech in quiet, there was no interaction ($p \geq .05$) between amplification type and speech level for any dimension, although the interaction for loudness was marginally significant ($F[4,68]=2.51$, $p=.050$). The trend for loudness was consistent with previous studies (e.g., Jenstad et al, 2000); namely, conversational or loud speech was louder with PC than with WDRC, and soft speech was softer with PC than with WDRC.

For overall impression, rating varied with speech level ($F[2,34]=11.30$, $p=.001$), with the highest rating for the conversational speech and the lowest rating for the loud speech. For loudness, the rating was highest for the loud speech and lowest for the soft speech ($F[2,34]=266.15$, $p<.001$), as expected. For pleasantness, the loud speech was least pleasant, with similar ratings for soft and conversational speech ($F[2,34]=13.92$, $p<.001$). For intelligibility, ratings were near maximum for conversational and loud speech, and lower for soft speech ($F[2,34]=9.98$, $p=.002$). Although not statistically significant, the direction of the ratings for PC versus WDRC was consistent with the expectation of distortion due to clipping. That is, listeners rated soft speech as similarly intelligible for both amplification types, but the mean rating for loud speech was lower for PC than for WDRC. There was no effect of amplification type for loudness ($F[2,34]=1.25$,

Table 4. Mean Speech Quality Ratings

	Stimulus	PC	CL	WDRC
Overall impression	50 dB SPL	6.8	6.8	6.9
	65 dB SPL	7.5	8.3	8.0
	80 dB SPL	4.7	5.1	5.8
	+6 SNR	6.7	6.1	6.6
	+10 SNR	7.2	6.8	7.3
Loudness	50 dB SPL	2.2	2.4	2.5
	65 dB SPL	6.5	5.6	5.6
	80 dB SPL	9.4	9.3	9.2
	+6 SNR	4.7	5.8	4.6
	+10 SNR	4.8	4.6	4.3
Pleasantness	50 dB SPL	6.9	7.1	6.7
	65 dB SPL	7.4	7.4	8.1
	80 dB SPL	4.2	3.9	4.3
	+6 SNR	6.2	6.1	6.3
	+10 SNR	6.9	6.8	7.3
Intelligibility	50 dB SPL	7.9	7.8	7.9
	65 dB SPL	9.7	9.3	9.3
	80 dB SPL	9.1	9.2	9.7
	+6 SNR	8.3	8.4	8.0
	+10 SNR	9.1	8.9	8.7

Note: Except for loudness, 1 = least favorable and 10 = most favorable. For loudness, 5 = most favorable.

p=.299), overall impression (F[2,34]=1.40, p = .260), pleasantness (F[2,34]=.30, p=.740), or intelligibility (F[2,34]=.35, p=.707).

For speech in noise, rating scores for each of the four quality dimensions were analyzed using two-way repeated measures ANOVA. In each case, the within-subject factors were amplification type and noise SNR (+6, and +10). The interaction between amplification and SNR was not significant for any quality dimension (p>.05). The +6 SNR condition was rated as less pleasant (F[1,17]=7.87, p=.012), less intelligible (F[1,17]=6.48, p=.021), louder (F[1,17]=5.04, p=.038), and of poorer overall quality (F[1,17]=5.30, p=.034) than the +10 SNR condition. Consistent with previous work (Souza et al, 2006), none of the amplification strategies as implemented without digital noise reduction or directional microphones offered an advantage when listening in noise (p>.05).

Music Quality

For music (Table 5), mean ratings were analyzed using a two-factor repeated measure ANOVA. Within-subject factors were amplification type and music genre. For each dimension, there was no interaction between music genre and amplification type (p>.05). Classical music was rated as preferred overall (F[1,17]=14.70, p=.001), less loud (F[1,17]=13.53, p=.002), less sharp (F[1,17]=6.46, p=.021), and more pleasant (F[1,17]=12.08, p=.003) than the popular selection. The genres were rated as equally full (F[1,17]=1.27, p=.276). There was no effect of amplification type on music loudness (F[2,34]=2.36, p=.128), sharpness (F[2,34]=1.03, p=.348), or fullness (F[2,34]=.68, p=.513). There was a significant effect of amplification type on pleasantness (F[2,34]=6.19, p=.005) and a marginally significant effect of amplification type on overall impression (F[2,34]=3.18, p=.054). In each case, ratings were higher for WDRC compared to the linear amplification strategies.

Contribution of Specific Dimensions to Overall Impression

Bivariate correlation analysis was used to evaluate the relationship between each of the rated dimensions of quality and overall impression ratings for each condition and input level. For speech (Table 6), the contribution of specific dimensions to overall impression varied with input level. For soft speech, both intelligibility and pleasantness were correlated with overall impression. For loud speech, intelligibility was not related to overall impression, presumably because intelligibility was maximized with mean scores greater than 9 for all amplification conditions.

Table 5. Mean Music Quality Ratings (n=18)

	Stimulus	PC	CL	WDRC
Overall impression	Classical	6.7	6.4	6.9
	Popular	5.3	5.6	6.4
Loudness	Classical	5.6	5.9	5.3
	Popular	6.7	6.9	5.9
Sharpness	Classical	5.8	5.3	5.4
	Popular	6.3	6.7	5.7
Fullness	Classical	6.0	6.4	6.2
	Popular	5.6	5.6	6.2
Pleasantness	Classical	6.4	6.2	7.0
	Popular	5.0	5.2	6.4

Table 6. Correlation Coefficients for the Correlation between Overall Impression and Sub-dimensions of Speech Quality

Overall Impression Sub-dimension (soft speech)	PC	CL	WDRC
Loudness	0.12	0.14	-0.2
Intelligibility	0.81**	0.66**	0.70**
Pleasantness	0.88**	0.75**	0.59*

Overall Impression Sub-dimension (conversational)	PC	CL	WDRC
Loudness	0.25	0.35	-0.02
Intelligibility	0.15	0.31	0.2
Pleasantness	0.61**	0.16	0.37

Overall Impression Sub-dimension (loud speech)	PC	CL	WDRC
Loudness	0.004	-0.08	0.09
Intelligibility	0.41	0.15	0.25
Pleasantness	0.85**	0.47*	0.63**

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Table 7. Correlation Coefficients for the Correlation between Overall Impression and Sub-dimensions of Music Quality

Overall Impression Sub-dimension (classical music)	PC	CL	WDRC
Loudness	-0.23	0.13	0.03
Sharpness	-0.09	-0.12	-0.02
Fullness	0.63**	0.55*	0.79**
Pleasantness	0.97**	0.85**	0.85**

Overall Impression Sub-dimension (vocal music)	PC	CL	WDRC
Loudness	0.34	-0.19	0.17
Sharpness	0.19	-0.24	0.13
Fullness	0.86**	0.85**	0.79**
Pleasantness	0.85**	0.89**	0.98**

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Thus, listener ratings of overall impression for loud speech were most strongly related to perceived pleasantness. For music (Table 7), fullness and pleasantness were most closely related to overall impression.

Speech Recognition

Mean HINT signal-to-noise ratio (SNR) scores were similar across the three amplification strategies: PC (-2.58), CL (-2.85), and WDRC (-3.01). HINT scores were analyzed using one-way repeated measures ANOVA, with amplification type as the within-subject factor. HINT scores across all subjects were not significantly different ($F[2,34]=.20$, p = .817) across the three amplification strategies.

DISCUSSION

For 80 dB SPL speech in quiet, there was a preference for wide dynamic range compression over linear amplification. Speech quality ratings were similar for 50 and 65 dB SPL. The preference for WDRC at high speech levels was consistent with previous reports (Jenstad et al, 2000; Kam and Wong, 1999; Larson et al, 2000; Lunner et al, 1998; Souza et al., 2002). For speech in noise, listeners rated the +10 SNR higher than +6 SNR, but ratings were equivalent for PC, CL, and WDRC. Omnidirectional WDRC amplification does not confer any acoustic advantage in noise (Souza et al, 2006), and, with minor exceptions, cannot improve speech recognition in noise either (for a review, see Souza, 2002). The latter point was reiterated in this study by the statistically identical HINT scores across amplification conditions. Thus, our listener's perceptions were consistent with both acoustic and intelligibility studies.

For music, there was a marginally significant effect of amplification type, with slightly higher ratings for the WDRC condition. Listeners also stated a strong preference for classical over vocal music. Chasin and Russo (2004) noted that unlike the redundancy noted in speech, musical genres differ in their frequency composition and perceptual emphasis. In addition, they noted that music has a 6-8 dB higher crest factor than speech. In this case, spectra of both musical renditions revealed a difference in the amplitude and frequency distribution of energy. Five second time segments of each piece are shown in Figure 2, and the long-term spectra (averaged over two minutes) are shown in Figure 3. The classical piece had a wider intensity range, higher crest factor (15 dB), and less high frequency energy, while the vocal piece had a smaller intensity range, lower crest factor (10 dB), and greater high frequency energy. The reason behind the difference in loudness ratings (with the popular piece rated as louder for all amplification types) is also suggested by Figure 2; although peak levels were similar, the popular piece had more periods of higher-level signal compared to the classical piece.

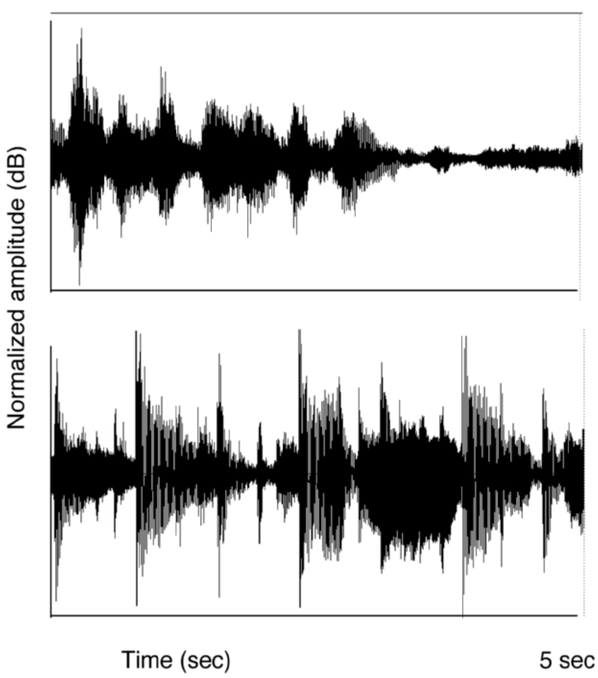

Figure 2. Five-second time segments of classical (top panel) and popular (lower panel) music.

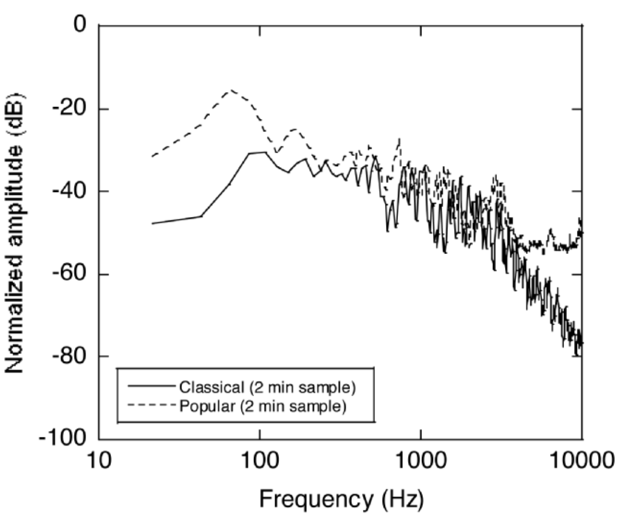

Figure 3. Frequency spectra (2 minute sample) for the classical (solid line) and popular (dashed line) music.

The difference in long-term spectra (Figure 3) may have been the source of the lower sharpness ratings for the classical piece; typically, "sharpness" is thought to be related to the level of high-frequency energy (Jenstad et al., 2003). Franks (1982) noted that although listeners with hearing loss prefer high-frequency emphasis for speech, they dislike the quality of high frequencies (4 kHz and beyond) in a musical context. Thus, listener's preference for classical music may be due to the reduced high-frequency energy in the spectra of the classical piece. Of course, it may also be the case that our listeners, aged 70 years on average, simply preferred classical music, even though we took care to select a popular piece that we considered to be age-appropriate.

Chasin and Russo (2004) also suggested that WDRC with low compression thresholds and longer release times may better accommodate the amplification requirements for music. That hypothesis was supported by the present data, which showed that WDRC was preferred over linear (PC or CL) amplification for some aspects of music quality, such as pleasantness and overall impression.

With regard to speech, some discrepancies still exist in regards to the best predictor of overall preference. Preminger and Van Tasell (1995b) showed correlations between pleasantness and overall impression for speech consistent with the present data. Other investigators have shown that clarity (Balfour and Hawkins, 1992; Gabrielsson et al, 1988) may influence listener's overall speech preference ratings. Although we did not ask listeners to rate "clarity," ratings of "intelligibility" were high for all speech levels and amplification conditions. In contrast to

Table 8. Mean HINT and Rating Scores for Experienced and New Hearing Aid Users

	Experienced Hearing Aid Users			New Hearing Aid Users		
Listening condition	PC	CL	WDRC	PC	CL	WDRC
HINT	-4.26	-3.97	-3.39	-0.9	-1.73	-2.62
50 dB SPL	6.8	6.9	7.4	6.8	6.9	6.3
65 dB SPL	6.6	8.3	8	8.4	8.2	8
80 dB SPL	3.4	3.3	5	6	6.9	6.7
6 SNR	6.2	6	6.5	7	6	6.7
10 SNR	7.4	6.9	7.7	6.9	6.8	6.9
Mean rating for speech	6.08	6.28	6.92	7.02	6.96	6.92
Classical music	6	6	6.1	7.4	6.8	7.7
Vocal music	5.3	4.9	6.2	5.3	6.2	6.7
Mean rating for music	5.65	5.45	6.15	6.35	6.5	7.2

previous work, perhaps these advanced-technology aids that had less distortion and a wider bandwidth than older aids simply eliminated poor intelligibility as a factor. We did not find that loudness influenced overall impression, although other studies have shown that perceived loudness influences acceptance of amplification (e.g., Leijon et al, 1991).

To our knowledge, this is the first study to parse out specific attributes that contribute to music quality. For music, fullness and pleasantness directly correlated with overall impression ratings; loudness and sharpness did not. This contrasts with clinical convention when adjusting a hearing aid for speech, where complaints about loudness or sharp sound quality are cited as common reasons for rejection of the aid (Kochkin, 2000b). In other words, the perceptual quality dimensions commonly used for speech quality rating may not yield useful information for guiding hearing aid adjustments for the hearing impaired listener who desires to listen to other signals.

Finally, previous studies that evaluated subjective performance for different signal processing strategies for speech in quiet and noise at various input levels did not make a direct comparison between experienced and new hearing aid users. We know that new users are different; they have been reported to prefer less overall gain (e.g., Marriage et al, 2004); receive more objective (Cox et al, 1991; Cox and Alexander 1992) and subjective benefit (Cox and Alexander, 1992); and have different signal processing preferences (Ricketts and Bentler, 1992) compared to experienced users. Although the subject number was too small for a comprehensive comparison, the new and experienced users had similar (although not statistically identical) ages and audiograms. Therefore, it is interesting to consider a general comparison between these two groups. In this study, HINT and quality rating scores were equivalent between the experienced and new hearing aids users who participated in this study. These results (Table 8) suggest that signal processing preference does not depend on hearing aid experience.

CONCLUSIONS

The data presented in this study extend the work of previous investigators by providing a direct comparison of three commonly used amplification strategies for speech and music stimuli. Information on the relationship between specific dimensions of speech quality and overall preference for a given amplification strategy revealed that intensity level and signal-to-noise ratio influenced ratings of each dimension for all three hearing aid circuits. For music, pleasantness and fullness had the greatest influence on overall preference. These findings may offer a streamlined way to obtain listener's preference when adjusting a hearing aid in a clinical setting. Thus, different amplification strategies may be suited to individual listening needs and desires. Such information could be obtained informally or via a prioritized listing metric such as the COSI. Equivalent HINT scores suggested that subjective amplification preference will not compromise objective performance. Results also suggest that these findings are not dictated by the amount of previous hearing aid experience.

Acknowledgments. This project was supported by grant R01 DC006014 and by Phonak Hearing Instruments. The authors gratefully acknowledge the assistance of Caitlin Rawn in collecting the HINT list equivalency data.

NOTES

1. Because published reports of HINT list equivalency used different speaker configurations, list equivalency was confirmed for this study using the 0/180 degree azimuth configuration in a separate group of ten listeners (mean age 21 years) with normal hearing. For each listener, HINT threshold was measured for each of the 25 lists, presented in random order. Mean list thresholds ranged from -10.2 dB to -12.9 dB. A one-way, repeated-measures ANOVA confirmed that the HINT lists were statistically equivalent ($F[24,216]=1.16$, $p=0.279$).

2. For each reported analysis, Mauchly's test (Mauchly, 1940) was evaluated and the Greenhouse-Geiser adjusted values were used if the assumption of sphericity was violated.

REFERENCES

Balfour PB, Hawkins DB. (1992) A comparison of sound quality judgments for monaural and binaural hearing aid processed stimuli. *Ear Hear* 13(5):331–339.

Byrne D, Dillon H. (1986) The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear Hear* 12:47–54.

Chasin M. (2003) Five differences between speech and music for hearing aids. *Hear J* 56:7.

Chasin M, Russo FA. (2004) Hearing aids and music. *Trends Amplif* 8(2):35–47.

Compton-Conley CL, Neuman AC, Killion MC, Levitt, H. (2004) Performance of directional microphones for hearing aids: real-world versus simulation. *J Am Acad Audiol* 15(6):440–455.

Cox RM, Alexander GC. (1992) Maturation of hearing aid benefit: objective and subjective measurements. *Ear Hear* 13(3):131–141.

Cox RM, Alexander GC, Rivera IM. (1991) Comparison of objective and subjective measures of speech intelligibility in elderly hearing-impaired listeners. *J Speech Hear Res* 34:904–915.

Cox RM, McDaniel DM. (1984) Intelligibility ratings of continuous discourse: application to hearing aid selection. *J Acoust Soc Am* 76:758–766.

Cox RM, McDaniel DM. (1989) Development of the speech intelligibility rating (SIR) test for hearing aid comparison. *J Speech Hear Res* 32:347–352.

Feldmann H, Kumpf W. (1988) Listening to music in hearing loss with and without a hearing aid. *Laryngorhinootogie* 67:489–497.

Gabrielsson A, Schenkman BN, Hagerman B. (1988) The effects of different frequency responses on sound quality judgments and speech intelligibility. *J Speech Hear Res* 31:166–177.

Hawkins DB, Walden BE, Montgomery A, Prosek RA. (1987) Description and validation of an LDL procedure designed to select SSPL90. *Ear Hear* 8(3):162–169.

Jenstad LM, Pumford J, Seewald RC, Cornelisse LE. (2000) Comparison of linear gain and wide dynamic range compression hearing aid circuits II: aided loudness measures. *Ear Hear* 21(1):32–44.

Jenstad LM, Van Tasell DJ, Ewert C. (2003) Hearing aid troubleshooting based on patients' descriptions. *J Am Acad Audiol* 14(7):347–360.

Jerger J. (1970) Clinical experience with impedance audiometry. *Arch Otolaryngol* 92:311–324.

Kam AC, Wong LL. (1999) Comparison of performance with wide dynamic range compression and linear amplification. *J Am Acad Audiol* 10:445–457.

Kochkin S. (2000a) MarkeTrak V: consumer satisfaction revisited. *Hear J* 53:38–55.

Kochkin S. (2000b) "Why my hearing aids are in the drawer": the consumer's perspective. *Hear J* 53:34–41.

Larson VD, Williams DW, Henderson WG, et al. (2000) Efficacy of 3 commonly used hearing aid circuits: a crossover trial. *J Am Med Assoc* 284(14):1806–1813.

Leijon A, Lindkvist A, Ringdahl A, Israelsson B. (1991) Sound quality and speech reception for prescribed hearing aid responses. *Ear Hear* 12:251–260.

Lunner T, Hellgren J, Arlinger S, Elberling C. (1997) A digital filterbank hearing aid: three digital signal procession algorithms-user preference and performance. *Ear Hear* 18(5):373–387.

Lunner T, Hellgren J, Arlinger S, Elberling C. (1998) Non-linear signal processing in digital hearing aids. *Scand Audiol Suppl* 49:40–49.

Marriage J, Moore BC, Alcantara JI. (2004) Comparison of three procedures for initial fitting of compression hearing aids. III. Inexperienced versus experienced users. *Int J Audiol* 43(4):198–210.

Mauchly JW. (1940) Significance test for sphericity of a normal n-variate distribution. *Ann Math Stat* 11:204–209.

Mueller HG. (1992) Insertion gain measurements. In: Mueller HG, Hawkins DB, Northern JL. *Probe Microphone Measurements: Hearing Aid Selection and Assessment.* San Diego: Singular Publishing Group, 113–144.

Nilsson M, Soli SD, Sullivan JA. (1994) Development of the Hearing in Noise Test for the measurement of speech reception in quiet and noise. *J Acoust Soc Am* 95:1085–1099.

Preminger JE, Van Tasell D. (1995a) Measurement of speech quality as a tool to optimize the fitting of a hearing aid. *J Speech Hear Res* 38:726–736.

Preminger JE, Van Tasell D. (1995b) Quantifying the relationship between speech quality and speech intelligibility. *J Speech Hear Res* 38:714–725.

Punch JL. (1978) Quality judgments of hearing aid-processed speech and music for normal and otopathologic listeners. *J Am Audiol Soc* 3:179–188.

Ricketts TA, Bentler RA. (1992) Comparison of two digitally programmable hearing aids. *J Am Acad Audiol 3(2):101–112.*

Souza P. (2002) Effects of compression on speech acoustics, intelligibility and sound quality. *Trends Amplif* 6(4):131–165.

Souza P, Jenstad L, Boike K. (2006) Measuring the acoustic effects of compression amplification on speech in noise. *J Acoust Soc Am* 119:41–44.

Souza P, Yueh B, McDowell J, Collins M, Loovis C, Boike K, Deyo R. (2002) Sensitivity of self-assessment questionnaires to differences in hearing aid technology. Poster presented at the International Hearing Aid Conference, Lake Tahoe, CA.

Van Buuren RA, Festen JM, Houtgast T. (1996) Peaks in the frequency response of hearing aids: evaluation of the effects on speech intelligibility and sound quality. *J Speech Hear Res* 39(2):239–250.

Van Buuren RA, Festen JM, Houtgast T. (1999) Compression and expansion of the temporal envelope: evaluation of speech intelligibility and sound quality. *J Acoust Soc Am* 105:2903–2913.

Wilson RH, Zizz CA, Shanks J, Causey GD. (1990) Normative data in quiet, broad-band noise, and competing message for northwestern university auditory test NU-6 by a female speaker. *J Speech Hear Disord* 55(4):771–778.

Appendix A.

Dimensions of Quality Ratings for Speech

OVERALL IMPRESSION
Considering everything that you have heard, what do you think about the sound?

LOUDNESS
How loud, strong or forceful is the sound? The opposite of loud is soft, weak or timid/faint.

INTELLIGIBILITY
How clear is the speech or what percent of the speech do you understand? The opposite of intelligible is extremely hard to understand and unclear.

PLEASANTNESS
How pleasing is the tonal quality of the sound? The opposite of pleasant is unpleasant.

Appendix B.

Dimensions of Quality Ratings for Music

LOUD
How loud, strong or forceful is the sound?
The opposite of loud is soft, weak or timid/faint.

SHARPNESS
How hard, keen or shrill is the sound?
The opposite of sharp is gentle and soft.

FULLNESS
How full is the sound?
The opposite of full is thin.

PLEASANTNESS
How pleasing is the tonal quality of the music?
The opposite of pleasant is unpleasant.

OVERALL IMPRESSION
Considering everything that you have heard, what do you think about the sound?