TRANSMISSION OF PHONETIC INFORMATION IN LINEAR FREQUENCY TRANSPOSITION

Petri Korhonen, M.Sc. (Tech.),
Widex Office of Research in Clinical Amplification (ORCA), Widex USA

INTRODUCTION

The transmission of phonetic information in phoneme identification when using a new linear frequency transposition method was studied. Seven normal-hearing young test subjects with simulated high-frequency hearing loss at and above 1600Hz were trained and tested in consonant and vowel identification task using CVC and VCV nonsense syllables.

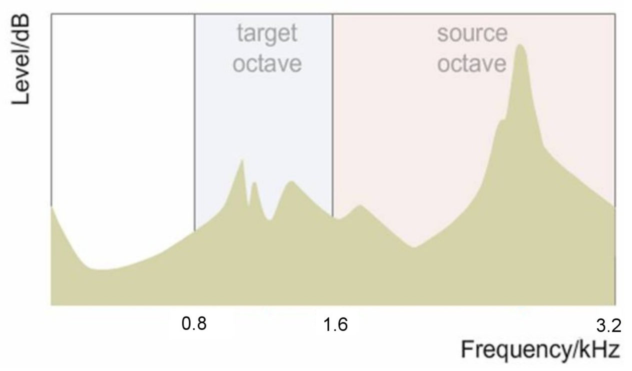
The Widex Inteo™ hearing aid uses a linear frequency transposition algorithm called the Audibility Extender (AE). This algorithm is designed to help people with severe-to-profound hearing loss in the high frequencies hear the high-frequency sounds by transposing those sounds to lower frequencies, where the hearing loss can be compensated with conventional acoustic amplification.

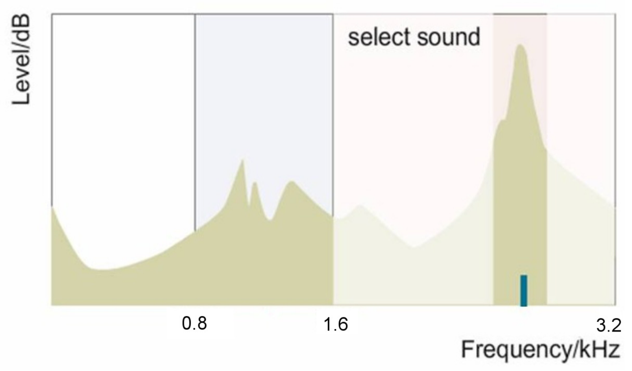
The frequency transposition algorithm creates new acoustic representations for high-frequency sounds which the listener may not have heard before. The usability of the new acoustic cues provided depends potentially on several factors, including the degree and the configuration of the hearing loss, the appropriateness of the amplification after transposition, and cognition and central processing, including reorganization of cortical representation through learning/training and motivation. To address the fundamental utility of frequency transposition, we want to minimize the confounding variability that may result from the interaction of these factors. Limiting or reducing the influence of confounding variables allows us to discern if frequency transposition provides usable cues that may benefit listeners. This was achieved by using normal hearing test subjects with identical simulated hearing loss across all subjects. Using NH subjects does not, however, allow us to determine if this algorithm would benefit individuals with hearing losses.

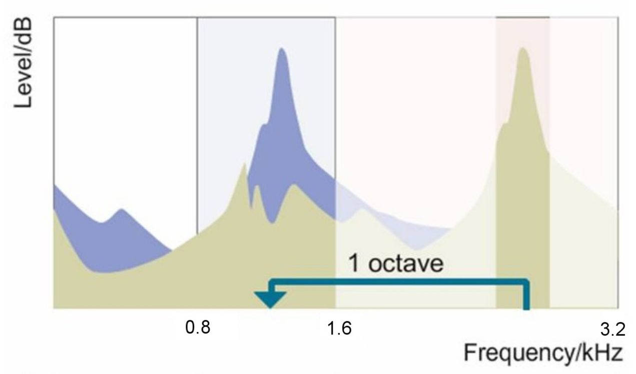
In earlier studies we have demonstrated the potential usability of these new acoustic cues in voiceless phoneme recognition (Korhonen 2007) using normal hearing subjects with simulated hearing losses. The current study extends the earlier study to cover all American English consonants.

AUDIBILITY EXTENDER (AE)

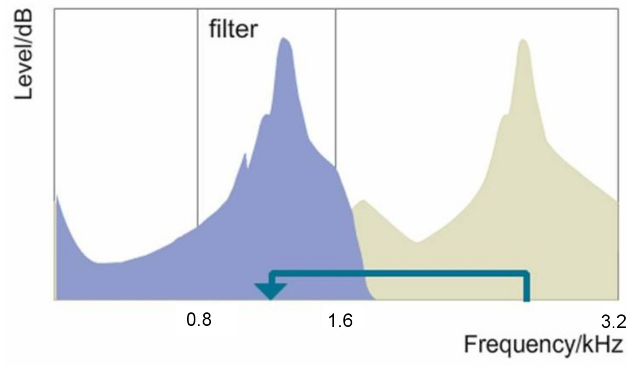
- The AE algorithm detects the most prominent peak located at the source octave, which is defined as the octave immediately above a user programmable start frequency.
- The frequency components above the start frequency are transposed. Those below start frequency are left unprocessed.
- Start frequency is set based on the location of the slope of the hearing loss.







- The largest peak is lowered downward by one octave and the other frequencies in the source octave are lowered linearly by the same amount.
- The transposed signal is then band-pass filtered around the transposed peak with a one octave bandwidth to limit the masking effects.



- Finally, the transposed sounds are amplified accordingly and mixed with the original signal as the final output (Andersen, 2006).

SUBJECTS AND STIMULI

- A total of seven normal hearing native English speakers ages 18-25 participated in the study (6 females, 1 male). Using normal hearing subjects allowed us to simulate an identical hearing loss across subjects and to minimize potential variability in cognitive functioning among subjects.
- Stimuli consisted of 22 consonants of American English: /m/, /n/, /j/, /b/, /d/, /g/, /v/, /w/, /ɹ/, /l/, /z/, / dʒ /, /ð/, /θ/, /f/, /ʒ /, /s/, /ʃ/, /tʃ/, /p/, /t/, /k/. Each phoneme was presented in medial positions in nonsense VCV syllables with vowels /a/, /I/, and /u/. Different vowels were used in different sessions. The syllables were spoken by a female native English speaker, and recorded (f_s=44.1kHz) in low noise level ($L_{A,slow}$ = 22dB) double walled audiometric test booth.
- After the initial recordings, the stimuli were played at 68dB SPL through a single loudspeaker placed 1 m in front of an Inteo IN-9 hearing aid connected to a 2-CC coupler. The output of the hearing aid in two different hearing aid programs was recorded. One was recorded with (AE On) and one without (AE Off) frequency transposition. The start frequency of the transposition was set to 1600Hz. Source region of the frequency transposition included sounds at two octave bands above the start frequency. In both programs the in-situ thresholds (sensogram) used to specify hearing aid gain was set to 20dB at and below 1000Hz, 50dB at 1250Hz, 70dB at 1600Hz, 90dB at 2000Hz and higher.
- To simulate hearing loss, the output recordings of the hearing aid were low pass filtered with a 8196-point FFT-filter (Blackman) with attenuation of 0dB for frequencies 0-1550Hz, linearly increasing attenuations between 1550Hz and 1650Hz, so that at 1650Hz and above, the attenuation was 90dB.

METHODS

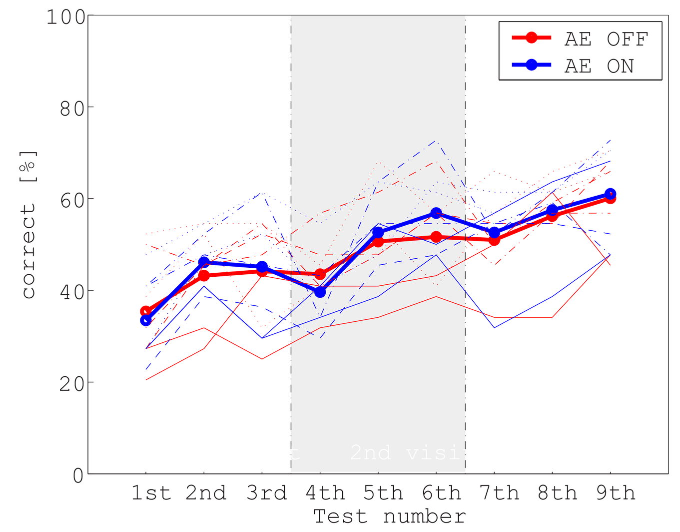
Identification test

- The tests took place in the training room at the US Widex Office of Research in Clinical Amplification (ORCA). The background noise in the office was $L_{A,slow}$ = 48 dB.
- The recorded stimuli were presented through headphones (Beyerdynamic DT 770) in random order. Subjects were allowed to adjust the presentation level to their desired level prior to testing (range L_A=76-79dB, mean 77dB). Both conditions (AE off & on) were tested separately.
- After each stimulus was presented the test subject was asked to identify the syllable by clicking on the corresponding button on the computer screen using a mouse. Each stimulus in the list was presented randomly twice. A one second pause was introduced after each answer before the next syllable was presented. The listeners had no option of repeating the stimulus.
- The phoneme identification test was conducted three times in three sessions which took place on separate dates within approximately 10 days. This resulted in combined of 9 identification tests for each subject.

Training

- Subjects participated in 10 minutes of self-paced training between each test presentation, but were not exposed to frequency transposed sounds between different sessions.
- The stimuli that were recorded with the "AE On" condition were used as the training materials.
- Buttons with written labels of all the syllables were shown on the computer screen. Subjects were allowed to listen to any of the syllables by clicking the corresponding buttons using a mouse.
- Subjects were informed of their errors at the end of each test so they could focus their attention to the most difficult sounds first.

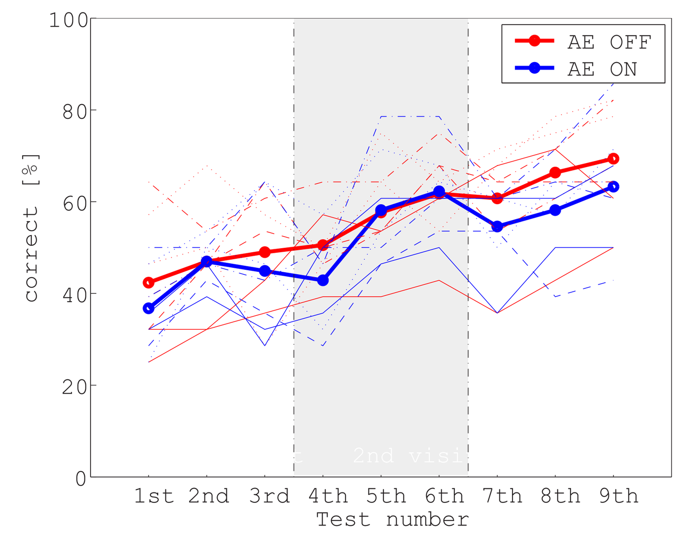
RESULTS



Identification for all consonants

ALL CONSONANTS

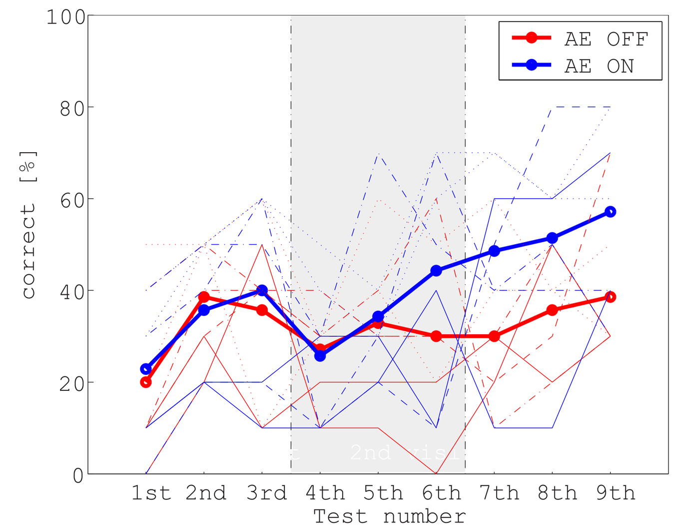
The best mean identification score (60.1% w/ AE off, 61.0% w/ AE on) was achieved at the 9th test. This difference in was not statistically significant ($p \leq 0.8402$). The mean score improved for both 'AE off' and 'AE on' conditions despite the training being done with AE-processed material. With 'AE on' condition, the mean score decreased between subsequent sessions, but always improved during the course of a given session above the best scores achieved at the previous sessions. Improvement between the 1st and 9th test was 24.7 percentage points (AE off)($p \leq 0.003$) and 27.6 percentage points (AE on)($p \leq 0.0045$).



Identification for voiced consonants

VOICED CONSONANTS

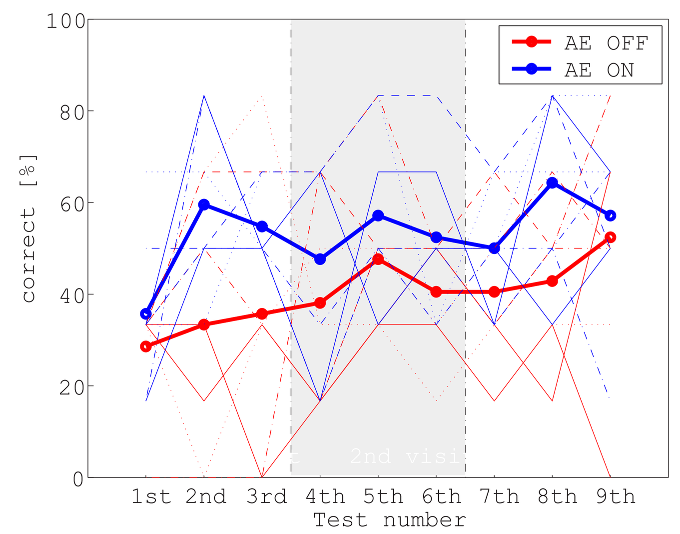
The best mean identification scores for voiced consonants (including voiced fricatives) were achieved at the 9th test. The mean scores were 69.4% (AE off) and 63.3% (AE on), respectively. This difference was not statistically significant (paired t-test: $p \leq 0.1683$). Performance with both conditions increased despite training being done using AE-processed speech sounds.



Identification scores for voiceless fricatives

UNVOICED FRICATIVES

Identification scores for voiceless fricatives for both 'AE on' and 'AE off' conditions show the mean identification scores at 9th test of 38.6% (AE off) and 57.1% (AE on). The difference is statistically significant ($p \leq 0.021$). The difference between these scores is 18.6 percentage points. The scores for two conditions were similar until the 6th test, after which the results for AE-processed surpassed those obtained with AE off. At the 6th test, the subjects had participated in 40 minutes of training.



Identification scores for voiceless stops

VOICELESS STOPS

The mean identification scores at the 9th test were 52.4% (AE off) and 57.1% (AE on), respectively. SPSS's general linear model analysis for repeated measurements for the scores from tests 2 through 9 shows that the difference between two conditions is statistically significant ($p<0.023$ w/ HA condition as a factor; $p<0.554$ w/ amount of training as a factor). The difference in averages from the 2nd through 9th tests was 14.0%. The difference between conditions was present immediately after first training session, unlike with voiceless fricatives.

RESULTS (CONT.)

SEQUENTIAL INFORMATION ANALYSIS

Sequential information analysis (SINFA) (Wang & Bilger, 1973) gives information about the transmitted information by iteratively identifying the phonetic feature for which the transmitted information (TI) is highest and keeping it constant at subsequent iterations. The results from the 9th test were analyzed for transmission of three features: voicing, manner and place of articulation. Mean transmitted information was 10% more with AE on condition (3.400 w/ AE off; 3.743 w/ AE on) ($p \leq 0.01$).

Voicing was transmitted equally well with both conditions (88.31, AE off vs. 88.90, AE on). The manner of articulation was transmitted slightly better with AE on (87.23 w/ AE off vs. 90.02 w/ AE on), but this difference was not statistically significant ($p \leq 0.5741$). The place of articulation was transmitted most poorly with both conditions, but was better transmitted with 'AE on' (41.22 w/ AE off vs. 52.82 w/ AE on)($p \leq 0.086$).

Percent of information transmitted for different features in rationalized arcsine units (RAU)

	VOICING		MANNER		PLACE	
	AE off	*AE on*	*AE off*	*AE on*	*AE off*	*AE on*
S1	74.81	85.46	76.22	82.98	10.96	43.54
S2	97.71	118.37	87.68	93.41	51.93	63.15
S3	99.94	59.83	95.32	75.24	38.38	55.16
S4	72.39	76.44	86.36	86.36	51.93	39.70
S5	84.45	86.62	79.79	92.19	49.82	51.47
S6	102.17	77.20	98.99	98.80	35.21	58.80
S7	83.10	118.37	89.87	101.13	50.28	57.95
mean	**87.80**	**88.90**	**87.75**	**90.02**	**41.22**	**52.82**

DISCUSSION

An extension of the previous study (Korhonen 2007) was conducted to investigate the identification of American English phonemes when using linear frequency transposition. Using normal hearing test subjects with simulated hearing loss at and above 1600Hz showed that AE processing improved the identification of both voiceless fricatives and stops. The improvement was 18.6% for voiceless fricatives and 14.0% for voiceless stops. This improvement comes at the expense of a small reduction in the identification of voiced consonants. To address the relative importance of identification of different phoneme groups in the English language is beyond the scope of this study.

The subjects were not able to successfully gain an additional benefit from the new acoustic cues without first learning to associate the new sounds with corresponding phonemes. The benefit of AE processing was apparent only at later visits (>30 minutes of training). The material used in the training was the same as was used in the testing. This prevents us from generalizing the results further to prove that AE will always help in identification of fricatives and stop consonants. However, the results obtained in this study demonstrate that AE does produce new acoustic cues that listeners with a simulated hearing loss at and above 1600Hz can potentially be trained to utilize in speech perception.

The training had a significant effect on identification results with the 'AE on' condition, but also improved the identification scores with the 'AE off' condition, despite the training being conducted using AE processed material. The improvement with 'AE off' may be credited to procedural learning, but it should also be noted that the AE processing retains the original signal at lower frequencies. Therefore, if the acoustic cues used for a particular identification were originally mostly at low frequencies (as is the case with most voiced phonemes), the training with AE processed material may also help in the identification of 'AE off' material.

Sequential information analysis for the results from the 9th test showed that AE processing increased the total amount of transmitted information by 10%. Further analysis with different phonetic features did not reveal which of the three phonetic features (voicing, manner, place) was responsible for the increase in total information transmitted. However, out of these three features, the transmitted information about the place of articulation was increased the most in AE processing.

REFERENCES

Andersen, H. 2006. Audibility Extender – So the "Dead" (Region) May Hear, Integrated Signal Processing, A New Standard in Enhancing Hearing Aid Performance, Widex Inteo.

Korhonen, P. 2007. Effect of training on frequency transposition. American Auditory Society meeting, March 4-6, 2006, Scottsdale, Arizona.

Kuk, F., Korhonen, P., Peeters, H., Keenan, D., Jessen, A., Andersen, H. 2006. Linear frequency transposition: Extending the audibility of high frequency information. Hearing Review. 13(10): 42-48.

Studebaker, G. 1985. A "rationalized" arcsine transform. J Speech Hear Res, 28, 455-462.

Wang, M.D., Bilger, R.C. 1973. Consonant confusions in noise: A study of perceptual features. J. Acoust. Soc. Am., 54, 1248-1266.