Potential Mechanisms for Perception of Frequency-Lowered Speech

Joshua M. Alexander
Ph.D., CCC-A

Disclosures

• I have received past research support from
  – The National Institutes of Health
  – The Indiana Clinical and Translational Sciences Institute
  – The William Demant Holding Group

• I hold a patent on a novel form of frequency lowering

• I am a paid consultant for Creare, LLC on an NIH SBIR project to develop an open-source, master hearing aid

• I have received speaker honoraria from Signia, Oticon, Phonak, Starkey, and ReSound hearing aid companies

Acknowledgements

Joshua M. Alexander
**Agenda**

- **Key point:** frequency-lowering methods and settings that preserve the greatest amount of temporal modulation from the original speech at the level of the auditory periphery will yield the best outcomes for speech perception.

1. Big picture: who, what, and why of frequency lowering
2. Paradigm shift from the frequency domain to the time domain for developing a theoretical model that describes perception of frequency-lowered speech
3. Evaluate different models/mechanisms using the newest method of frequency lowering
4. Critically evaluate current clinical recommendations for selecting frequency lowering parameters
Frequency Lowering is Ubiquitous

(Alexander, 2016b)

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Feature Name</th>
<th>Frequency Lowering Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Widex</td>
<td>Audibility Extender</td>
<td>Transposition (static)</td>
</tr>
<tr>
<td></td>
<td>Enhanced Audibility Extender</td>
<td>Transposition (adaptive)</td>
</tr>
<tr>
<td></td>
<td>Enhanced Audibility Extender</td>
<td>Transposition (adaptive)</td>
</tr>
<tr>
<td>Phonak</td>
<td>SoundRecover</td>
<td>Compression (static)</td>
</tr>
<tr>
<td></td>
<td>SoundRecover</td>
<td>Compression (static)</td>
</tr>
<tr>
<td>Starkey</td>
<td>Spectral iQ</td>
<td>Spectral Envelope Warping</td>
</tr>
<tr>
<td>Signia</td>
<td>Frequency Compression</td>
<td>Compression</td>
</tr>
<tr>
<td>ReSound</td>
<td>Sound Shaper</td>
<td>Proportional Compression</td>
</tr>
<tr>
<td>Oticon</td>
<td>Speech Rescue</td>
<td>Multilayered Transposition</td>
</tr>
</tbody>
</table>

* Also offered by Unitron as “Frequency Compression” and by Bernafon as “Frequency Composition” and by Sonic as “Frequency Transfer”
* Also offered by Unitron as “Frequency Compression” and by Bernafon as “Frequency Composition” and by Sonic as “Frequency Transfer”
* Also offered by Rexton as “Bandwidth Compression”
* Also offered by Beltone as “Sound Shifter” and by Interton as “Frequency Shifter”

Nonlinear Frequency Compression

Before Lowering

* Formant transitions
After Lowering

Nonlinear Frequency Compression

Control

Consonant (VCV) Identification

Alexander (in revision)

Benefits

Side Effects

Alexander 2016a

(n = 14 H subjects)

Benefits /s/ Identification Input Bandwidth (Hz) /s/ Identification

Side Effects Vowel Identification Input Bandwidth (Hz) Consonant (VCV) Identification

NFC

* Formant transitions (now flattened)*

* 50 % point of improvement
Potential Side Effects
Formant alteration, vowel reduction with a low cut off (start)

Potential Benefits and Side

Paradigm Shift

• From frequency domain to time domain
  • Some transposition and all compression methods adhere to an untested assumption that spectral features of high-frequency speech cues need to be preserved or replicated after lowering
  • Previous attempts to ‘optimize’ in terms of frequency
    • Input bandwidth (i.e., upper freq. limit)
    • Frequency separation (i.e., between key speech contrasts, namely [s] and [ʃ])
  • Preservation of features in the temporal modulation is not a new concept with respect to freq. lowering
    • It was the very premise behind most early methods of frequency transposition, especially vocoding methods
Recommended Clinical Procedure

- Scollie recommends a 2-step procedure using calibrated [s] and [ʃ] (e.g., Scollie et al., 2016)
  1. Make high-frequency ‘edge’ of [s] audible (≤ MAOF)
  2. Make high-frequency edge of [s] and [ʃ] non-overlapping, if possible
    - Phonak also recommends at least 1/3-octave (=1 ERB) separation between the low-frequency edges of [s] and [ʃ]

Untested Assumptions

1. [s]-[ʃ] discrimination is the same on the low-frequency edge as the high-frequency edge
2. [s]-[ʃ] edge frequency discrimination is constant across frequency and level
3. Maximizing [s]-[ʃ] edge frequency differences will maximize their discriminability
4. Maximizing [s]-[ʃ] discriminability will maximize perception of other fricatives and consonants

Unanswered Questions

1. What is the minimum frequency difference (just noticeable difference, JND) needed for normal-hearing listeners to tell two bands of noise apart?
2. How about for hearing-impaired listeners?
3. How does the addition of intervening sounds influence the JND for noise discrimination
Bandwidth Discrimination

- 12 normal-hearing (NH), college-aged listeners
- Filtered pink noise samples (level/ERB = constant)
  - 2 levels: 55 and 75 dB SLP/ERB
  - 150 msec (25-msec ramps) = duration of medial [s], [ʃ]
- 3-interval, forced choice task
  - For each trial, the noise samples were frozen
    - Listener had to identify the interval with the ‘oddball’
      - (a) wider LF edge, (b) wider HF edge, (c) both edges balanced

Bandwidth Discrimination

- 3 frequency ranges for the standard
  - Roughly correspond to average frequency and BW of lowered [s], [ʃ] for frequency-lowering settings
  - Appropriate for hearing losses that are (1) severely-profound (SP), (2) moderately-severe (MS), (3) mild-to-moderate (MM)
  - +/- 1 ERB rove on each trial (forced within trial comparisons)
  - 9 interleaved tracks: 3 freq ranges x {LF edge, HF edge, balanced} → tough!
- 2 down, 1 up steps (=70.7% correct)
  - \( \frac{1}{2} \) ERB (rev. 1-4), \( \frac{1}{2} \) ERB (rev. 5-8), \( \frac{1}{8} \) ERB (rev. 9-16)

Bandwidth Discrimination

- No main effect for level
- No main effect for [low, mid, high]
- Main effect for edge freq: \( LF > Balanced > HF \)

* r(16) = -0.62, p < 0.01
Auditory Nerve Model Simulations

- AN model simulations strongly suggest that HI subjects will have an opposite pattern of results as NH subjects.
- Indicates that discrimination of frequency-lowered frication using spectral separation alone will be difficult, especially on the HF edge, and as hearing loss increases.

Bandwidth Discrimination Summary

- **Tested Assumptions**
  1. Discrimination is the same on the low-frequency edge as the high-frequency edge.
     - **False**, discrimination is better on the high-freq edge.
  2. Discrimination is constant across freq & level.
     - **True** (for noise in isolation).

- **Answered Questions**
  1. Minimum freq. difference (JND) needed for normal-hearing listeners to tell two bands of noise apart?
     - ≈ 1.4 ERB (HF edge), ≈ 1.6 ERB (Balanced), ≈ 2.15 ERB (LF edge).
  2. How about for hearing-impaired listeners?
     - **??? Cannot use HF edge!**
  3. How does the addition of intervening sounds influence the JND for noise discrimination?
     - ↑ by ≈ 1.25 ERB, some effects of level and freq.

Untested Assumptions

1. [s]-[ʃ] discrimination is the same on the low-frequency edge as the high-frequency edge.
2. [s]-[ʃ] edge frequency discrimination is constant across frequency and level.
3. Maximizing [s]-[ʃ] edge frequency differences will maximize their discriminability.
4. Maximizing [s]-[ʃ] discriminability will maximize perception of other fricatives and consonants.
Adaptive Nonlinear Frequency Compression

Conventional NFC

Adaptive NFC (ANFC)
Full Bandwidth

"seven seals were stamped on great sheets"

Simulated Loss
(Filtering >3.5kHz)

Processed with ANFC (Stage 1)

 Mostly formants
(vowels, glides, liquids, stop transitions)

Stage 1
Source Region = 2.0 – 6.0 kHz
Destination Region = 2.0 – 3.5 kHz
Processed with ANFC (Stage 2)

Mostly frication
(fricatives, affricates, stop bursts)

Stage 2
Source Region = 2.0 – 11 kHz
Destination Region = 0.8 – 3.5 kHz

Processed with ANFC

Stage 1
Source Region = 2.0 – 6.0 kHz
Destination Region = 2.0 – 3.5 kHz

Stage 2
Source Region = 2.0 – 11 kHz
Destination Region = 0.8 – 3.5 kHz

Source Region for Stage 2 (lower cutoff)
Depends on Stage 1 (upper cutoff)

http://tinyurl.com/FLassist

Joshua M. Alexander
1. [s]-[ʃ] discrimination in words:
   - s-sh Confusion Test (Alexander, 201X)
   - 66 word pairs (132 total), 1 woman talker

2. Fricative discrimination (/iC/)
   - 7 fricatives, 3 women talkers

3. Consonant discrimination (/VCV/)
   - 20 consonants, 3 vowels, 1 man + 1 woman talker

4. Vowel discrimination (/hVD/)
   - 12 vowels, 2 men + 2 women

**Experimental Conditions**

- 3 groups of NH listeners (12 each)
  - Bandwidth limited to 2.1 kHz (SP), 3.3 kHz (MS), or 5.0 kHz (MM)
  - 8-9 different conditions processed in MATLAB that varied lower cut-off (FcL) and upper cut-off (FcU) with identical output bandwidth

**Results**

- Severe-to-Profound
  - MAOF = 2100 Hz

Increasing the max input beyond what is minimally necessary → blurring of the signal in time and frequency

- VCVs

Depends on phoneme:
  - /f/, /v/: ↑ % as FcU ↓
  - Small effect of FcL
Modulation Transfer Function Ratio (MTFR)

Potential Models/Mechanisms

• Maximizing input freq. or spectral separation often led to significantly lower speech recognition

• Narrower output BW was sometimes associated with higher recognition → Proxy for MTFR ???

• High correlations with the MTFR indices implicates the role of temporal modulation

Auditory Nerve Model Simulations

- 8-channel WDRC hearing aid simulator (e.g., Alexander, 2014)
- Zilany et al. (2014) model: Low SR with OHC/IHC integrity parameters automatically selected by the AN model for each audiogram (SP had complete OHC loss)
Tested Assumptions

3. Maximizing [s]-[ʃ] edge frequency differences will maximize their discriminability
   • FALSE, this tends to decrease discriminability
   • Preserving envelope modulation maximizes discriminability

4. Maximizing [s]-[ʃ] discriminability will maximize perception of other fricatives and consonants
   • TRUE for fricatives, mostly true for consonants as a class
   • Settings that best preserve envelope modulations for these sounds also do the same for other high-frequency sounds

Future Directions

• **Key point:** frequency-lowering methods and settings that preserve the greatest amount of temporal modulation from the original speech at the auditory periphery may yield the best outcomes for speech perception

• **Data collection**
  • Collect lab data on the experiments described today (ANFC) using hearing-impaired listeners
  • Later, will collect field data using clinical devices
  • Pilot an innovative variant of ANFC that I have developed
  • Started collecting data on an innovative variant of frequency transposition

Programming Software Adjustments

First slider controls destination region (cut-off 1 – max output)
Second slider controls cut-off 2 (and start of source region)
Setting "0" = max output
Given that we know:
1. The frequencies of the edges of the calibrated UWO [s] and [ʃ] files
2. The range of aided audibility (MAOF)
3. The relationship between input and output frequencies for the different ANFC settings
  • Cut-off 1, Cut-off 2, and Maximum Output frequencies

Then, we can compute:
1. The amount of separation between the high-frequency edges of the UWO [s] and [ʃ]
2. The amount of separation between the low-frequency edges of the UWO [s] and [ʃ]
3. The bandwidths of the lowered UWO [s] and [ʃ]
4. All of the above on a psychophysical scale that resembles normal cochlear filtering (ERB scale)
Research Papers


Invited Review Papers