

Evaluating Speech Segregation Performance in Human Listeners



Douglas S. Brungart

Air Force Research Laboratory



Introduction



- **Human listeners are generally good speech segregators**
 - **Robust in all kinds of environments**
 - **Performance far exceeds machine listeners**
- **But performance depends on many factors**
 - **How the target talker is designated**
 - **A priori information about the environment**
 - **What speech materials are used**
 - **Listener Familiarity and Training**
 - **Context**
- **Care must be taken to evaluate segregation in an environment that is representative of target application**



Target Talker Designation



In speech in noise tasks, listener knows what to listen to..

**-But when stimulus contains 2+ speech signals,
this may not be true**

**Very different results can be obtained depending on how
this designation is made....**



Target Talker Designation



Examples:

1) Known target talker...

- Lister is told to “listen for ‘Bob’ ”
- Requires significant training
- Does not allow target talker to change



Target Talker Designation



Examples:

2) Known target location....

- Lister is told to “listen in front”
- Valid only in spatially-separated conditions
- Does not allow target talker to be moved



Target Talker Designation



Examples:

3) Duration or Onset....

- Target phrase starts after the masker



Target Talker Designation



Examples:

- 4) Knowledge about content of target phrase
 - Embedded call sign (CRM)
 - “Topic” sentences (Freyman)
 - Listen for sentences about cooking



Target Talker Designation



Examples:

- 5) Knowledge about content of masking phrase
 - Transcript of content of masker phrases provided to listeners (Hawley et al.)



Target Talker Designation



Examples:

6) Visual cues

- Target talker is the visible talker (Driver)



Target Talker Designation



Examples:

7) Exhaustive Search

- Listener transcribes *all* talkers (Yost)



***A priori* information about environment**



Performance will vary depending on knowledge listener has about environment-

- 1) Where the target talker is/could be located**



***A priori* information about environment**



Examples:

- 2) Where the masking talkers are/could be located**



***A priori* information about environment**



Examples:

3) Voice characteristics of talkers

- Are individual talkers fixed in location?**



A priori information about environment



Examples:

4) Statistics of Environment

If target phrase changes talker and/or location,
how often does this happen?



A priori information about environment



Examples:

5) Stimulus characteristics

- Room characteristics
- Processing (applied to target/masker or both)



Speech Materials

Listener Familiarity



Selection of speech materials is also important...

Listener familiarity with materials is a serious issue...

- **Open-set sentences mimic real-world listening, but**
 - **Listener cannot know target or masker sentence**
 - **Sentences cannot be compared across conditions**
 - **New conditions cannot be tested with same subjects**
- **Context-free materials (CRM, MRT) can be re-used**
 - **Do not match real-world listening**



Speech Materials Training



Different materials require different training

Minimum Training

**Small response set
(CRM, MRT, PB50)**

PB250

Considerable Training

**Large Response Set
(PB1000)**

Open Set Words/Sentences

**Nonsense Syllables
(Transcription!?!)**



Speech Materials Context



Experimenters much choose what situation to model...

High Context- Possible spoken words constrained

(Everyday conversations?, Radio Discipline)

Emphasize effects of informational masking

Low Context- Listener has no idea of context

(Eavesdropping?)

Emphasize the effects of energetic masking

High Context

Low Context

CRM,PB-50

MRT,DRT,PB-250

PB-1000

**Meaningful
Sentences**

**Nonsense
Sentences**

**Nonsense
Syllables**