

# 「音声によるバリアフリー」に向けて

程島奈緒\*<sup>1</sup>

## Towards Barrier-free by Speech

by

Nao HODOSHIMA

(received on November 30, 2010 & accepted on December 24, 2010)

### Abstract

In loud and reverberant public spaces (e.g., train stations), speech announcements are often difficult to hear. This is especially true for elderly people, people with hearing impairments and non-native listeners for the language of the announcements. This paper introduces two approaches for improving speech intelligibility in public spaces: signal processing and speech production. The signal processing approach suppresses steady-state portions of speech before being radiated from loudspeakers. This approach reduces reverberant masking without degrading information in speech perception. The speech production approach uses the nature of human adaptation to speaking environments. For making announcements, speaking in noise or reverberation at appropriate playback level makes speakers modify their speaking styles, and such speech is more intelligible than speech spoken in a quiet environment, when they are heard in noise or reverberation. Listening tests with young and elderly participants showed that the two approaches statistically improved their speech intelligibility in several reverberant and noisy environments including simulated listening conditions and a lecture hall. The results imply that both approaches help to create barrier-free listening environments in public spaces. Finally, a possible application of the current findings would be in 1) developing the signal processing approach and 2) setting up recording conditions or instructing speakers how to effectively transmit messages to their audiences in public spaces where relatively higher speech intelligibility is required.

**Keywords:** Speech, Barrier-free, Intelligibility, Speech enhancement, Public spaces

**キーワード:** 音声, バリアフリー, 明瞭度, 音声強調, 公共空間

## 1. はじめに

音声は人がコミュニケーションを取るための重要な情報の一つである。人の音声情報処理の過程は speech chain<sup>1)</sup> というプロセスで表され、図1のように音響学的な段階(発話された音声の伝搬)、生理学的な段階(脳からの運動指令と発話、外耳から脳までの伝達)、言語学的な段階(脳での言語化や意味の理解)から構成される。この人の音声処理能力を基にして、電話、音声合成、音声認識、コーデック処理をはじめ、様々な音声技術が広範囲に及ぶ用途において開発されている。

音声コミュニケーションでは、発話者と聴取者間の距離が遠くなるにつれて伝送経路の影響をより受けやすくなる。例えば駅や空港などの公共空間で、スピーカから拡声された音声案内が聞き取りづらいと感じた経験はないだろうか。この原因の一つに雑音や残響(音が反射して響く現象)が挙げられる。雑音や残響は聴力が正常な若年者と比べて、高齢者、聴覚障害者、非母語話者(聞き取る言語が聴取者の母語ではない者)の明瞭度を大きく低下させる<sup>2)</sup>。

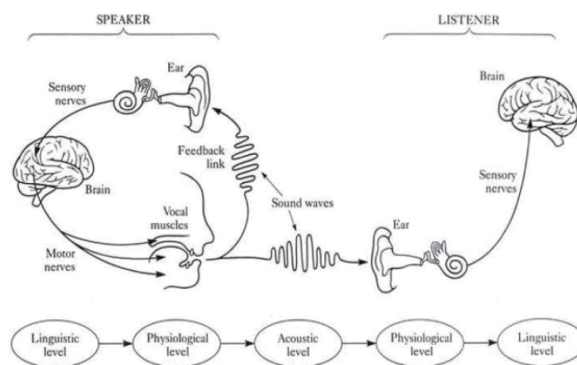


Fig.1 Speech chain. [Data is from the reference 1]

このうち日本では、高齢者の聞こえが特に重要な課題である。国内の高齢者人口は総人口の20%を超えており、2055年には2.5人に1人が高齢者になるとされている<sup>3)</sup>。一般的に内耳の細胞の機能は加齢と共に低下するため、高齢者の約2人に1人が加齢による難聴である。従って、今後日本では難聴者の著しい増加が見込まれる。また、高齢者の聞こえに関する調査<sup>4)</sup>では、80才代の55%が講演会や集会で聞き取れないと回答しており、聞き取れないことで分かったふりをしてしまう、外出がおっくうになるという回答が得られている。このことから、高齢者

\*1 情報通信学部情報メディア学科講師

が利用する空間での音声明瞭度は決して高いとはいえ、聞こえが個人の行動を左右することが示されている。

公共空間でのバリアフリー・ユニバーサルデザイン対策としては、2006年に障害者・高齢者の移動に関するバリアフリー新法<sup>5)</sup>の施行、2007年にバリアフリー整備ガイドライン<sup>6)</sup>の制定がされている。その結果、スロープやエスカレータの設置などの建築的なバリアフリーは緩和されてきている。その一方で音声伝達に関しては、バリアフリー整備ガイドライン<sup>6)</sup>に参考として述べられている限りで詳細な規定はない。その結果、例えば改札口付近に音案内を設置したが、案内が周囲の雑音に埋もれて聞こえない、音源の指向性が弱く改札口の方角が分からないなどの問題点が報告されており、音案内の設置が十分な効果を発揮していないとの指摘がある<sup>7)</sup>。

公共空間での音声案内は、空間の利用情報（例：鉄道の発着情報）のみならず、緊急時の放送にも使用される。後者に関しては、収容人数が300人以上の建築物では、火災などの緊急時にサイレン音だけではなく音声による放送を行うことが義務化されている<sup>8)</sup>。以上より、音声案内は正確な情報を伝達する手段として使用されているのにも関わらず、その情報を聞き取ることができない人達が存在する可能性があり、音声伝達の面でのバリアフリーは達成されているとはいえない。

そこで本研究では、公共空間で明瞭な音声案内を提供することで、音声のバリアフリーを実現することを目的とする。その実現に向けて、人の音声情報処理を取り入れた二つの手法を提案する。図2にその概念を示す。信号処理による手法を2章で、音声生成による手法を3章で述べる。

## 2. 提案手法1：信号処理

一つ目の手法では、公共空間の雑音や残響に頑健な音声案内を信号処理によって作成する。

### 2.1 信号処理

音声認識などで使用される一般的な雑音・残響除去は、雑音や残響が付加された観測信号から、ピッチ情報、調波構造、フォルマント構造などを利用して原音声の情報を復元する<sup>9,10)</sup>他)。本手法は従来法とは逆であり、公共空間に拡声する前のクリーンな音声に対して、雑音や残響の影響を予め軽減する信号処理を施す。この手法では信号処理は拡声前に行われるため、利用者は補聴機器を装着する必要がないという利点がある。また、この手法は既存の拡声システムの一部として導入することが可能なため建築的な改修が不要であり、低コストで明瞭度の改善を実現することができるという利点もある。

本手法では残響の特性と音声明瞭度の関係から、音声の定常部（母音の中央部のような定常的な音声区間）を抑圧する信号処理を用いる。定常部はスペクトルが変動する遷移部（閉鎖子音のような遷移を持つ音声区間）と比較して音声知覚にはそれほど重要ではないが、遷移部と比べて大きなエネルギーを持つ<sup>11)</sup>。残響下ではoverlap-masking<sup>12)</sup>という先行する音声区間が後続の音

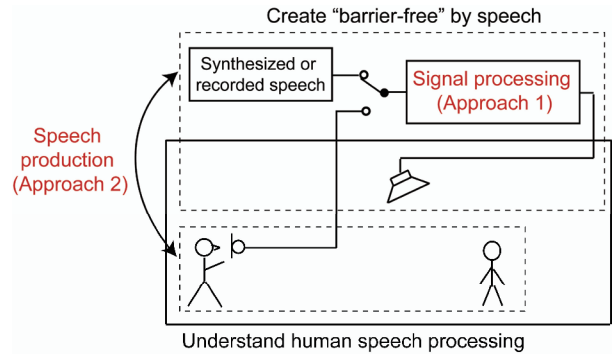


Fig. 2 Sketch of the proposed approaches for "barrier-free" by speech.

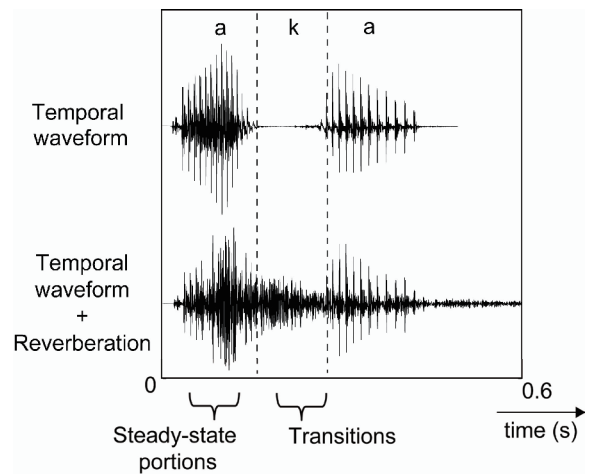


Fig. 3 Example of overlap-masking (top, temporal waveform of /aka/; bottom, reverberant one).

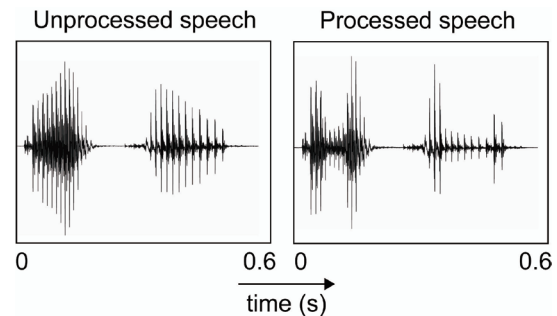


Fig. 4 Temporal waveforms of unprocessed and processed speech.

声区間をマスクする現象により、図3に示すように遷移部はエネルギーの大きい定常部に付加された残響の影響をより大きく受け、その結果として明瞭度が低下する。

そこで本手法は、音声の定常部を検出してそのエネルギーを予め抑圧してから拡声することで、音声知覚に必要な情報は最大限に保持しながら、overlap-maskingの影響を抑えて明瞭度の低下を軽減する。図4に処理前後の波形を示す。処理によって、エネルギーの大きい定常部が抑圧されていることが分かる。

## 2. 2 処理による明瞭度改善

本手法の効果を調べるため、聴取者に処理音声と提示して明瞭度を測定する心理実験（以下聴取実験）を行った。調査環境は公共空間を模擬した環境と講堂である。前者の環境では、原音声と処理音声にインパルス応答を畳み込み、条件によってはさらに雑音を付加した音声を、防音室内でヘッドホンから参加者に提示した。使用したインパルス応答は、複数の公共空間で測定されたもの、そのインパルス応答の残響時間（響きを表す物理指標の一つ）を変化させたもの、建築設計と拡声システムも含めて室内音響シミュレーションソフトウェアで作成したものを使用した。後者の環境では、講堂に参加者を配置し、原音声と処理音声を講堂内のスピーカから拡声して提示した。聴取実験で使用した音声は、ターゲット語を単音節とした文章である。

表1に処理による明瞭度の有意な改善（本稿では有意確率を5%とする）が得られた範囲をまとめる<sup>13-18</sup>他。また図5に、公共空間を模擬した環境と講堂における若年者と高齢者の結果を抜粋する。なお、処理のパラメータ

Table 1 Summary of conditions where the proposed approach improves speech intelligibility.

| Listeners   | Listening conditions  |
|---|---|
| Young normal hearing (total of 108 listeners)             | Simulated reverberant and noisy conditions                                  |
|   | Lecture room  |
| Elderly with/without hearing loss (total of 50 listeners) | Simulated rooms (including electroacoustical and architectural simulations) |
|   | Simulated reverberant conditions  |
|   | Lecture hall  |

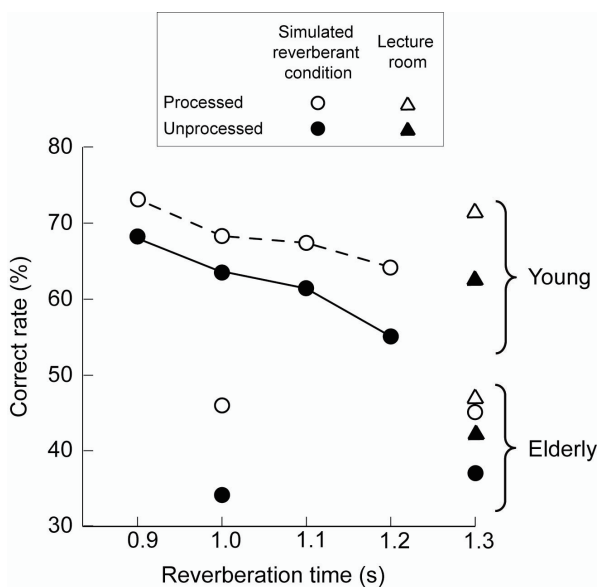


Fig. 5 Selection from results of listening tests for the approach 1. [Data is replotted from the references 14, 15 and 19]

は実験条件毎に最適化した。以上より、提案処理は特定の環境下において音声明瞭度を改善することが示された。

## 3. 提案方法2：音声生成

二つ目の手法では、人の音声生成の特徴を利用して、明瞭な音声案内を作成する発話環境の設定を行う。

### 3. 1 音声生成の特徴

人は正確なコミュニケーションを取るために、周囲の環境に応じて発話を変化させる。例えば雑音下での発話は、静かな環境での発話に比べて、時間や周波数に関する音響的特徴が変化する<sup>19, 20</sup>他。また、そのような雑音下で発話された音声の方が、静かな環境で発話された音声と比べて単語理解度が上昇する<sup>19, 20</sup>他。

その一方で、雑音とはマスキングの特性が異なる残響下ではこのような現象は調べられていない。公共空間では雑音に加えて残響も音声明瞭度を低下させる原因であるため、本手法ではまず残響下でも音響的特徴が変化するかを調査した。そして、残響下の発話の方が静かな環境で発話された音声よりも明瞭となるのかを調査した。

### 3. 2 音響分析

東京方言話者6名が、静か (Q)、雑音 (N)、残響 (R) 条件下で発話を行った。N は白色雑音を、R は残響時間が 3.6 s (R1) と 12.3 s (R2) のインパルス応答を使用した。発話内容は、親密度の統制が取れた単語をターゲット語とした文章である。録音は防音室で行い、雑音・残響条件では雑音または残響音（マイクに入力された音声にインパルス応答を遅延時間が数msで畳み込んだもの）をヘッドホンから発話者に提示した。発話者の耳元における提示音圧レベルは、4名には平均80dB、2名には発話者の口元の発話レベルに対して-22 dBとした。録音した音声には、雑音・残響のいずれも含まれていない。

4名の発話の音響分析結果<sup>21, 22</sup>を図6に示す。分析対象は基本周波数 (F0)、第1フォルマント (F1)、第2フォルマント (F2)、インテンシティ、子音対母音のインテンシティ比、時間長である。図6のアスタリスクは、発話

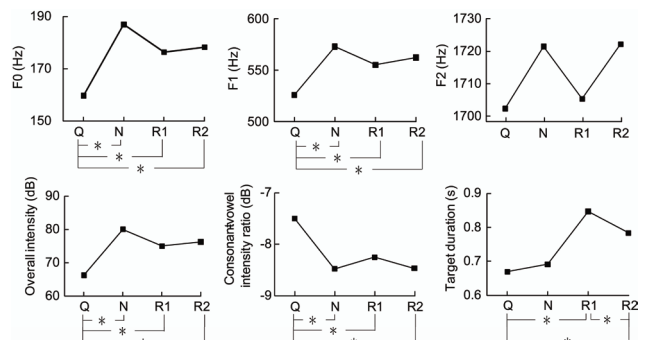


Fig. 6 Acoustic characteristics of speech spoken in quiet (Q), noise (N) and reverberation (R1 and R2). [Data is from the references 21 and 22]

条件間で有意な音響的特徴を示す。分析結果より、残響下での発話は静かな環境での発話と比べて音響的特徴の変化が観測された。さらに、雑音下と残響下では音響的特徴の変化は同一ではなく、発話の強調という面で同様の変化を示すもの (F0, F1, インテンシティ, 子音と母音のインテンシティ比) と、マスキングの違いにより異なる変化を示すもの (時間長) が確認された。さらに、発話者間の変動は雑音下よりも残響下で増加した。これは、定常的な雑音よりも、非定常かつ反射音による遅延が存在する残響音に対しては、発話者の経験や意識によっても適応の方法が異なる可能性が示された。

### 3. 3 発話環境による明瞭度改善

残響下で発話した音声 が明瞭になるかを調べるため、聴取実験を行った。実験で使用する刺激は、前節のQ, N, R1条件下で録音された2名の発話に、白色雑音 (信号対雑音比=-2 dB) の付加もしくはインパルス応答 (R1 と残響時間が2.6 sのR3) の畳み込みを行った音声で、表2に示す6条件とした。聴取者は健聴な若年の日本語母語話者32名であった。聴取実験は参加者毎に無響室内で行い、刺激はヘッドホンから提示した。

図7に聴取実験の結果を示す。雑音条件では、雑音下の発話 (N\_N) は静かな環境での発話 (Q\_N) よりも正解率が有意に高くなり、先行研究<sup>19, 20他)</sup>と一致した。雑音と残

Table 2 Listening conditions used in listening tests for the approach 2. [Data is replotted from the reference 23]

|                                      | Recording condition |     |       |
|--------------------------------------|---------------------|-----|-------|
|                                      | Q                   | N   | R1    |
| N<br>(signal to noise ratio = -2 dB) | Q_N                 | N_N |       |
| R1<br>(reverberation time = 3.6 s)   | Q_R1                |     | R1_R1 |
| R3<br>(reverberation time = 2.6 s)   | Q_R3                |     | R1_R3 |

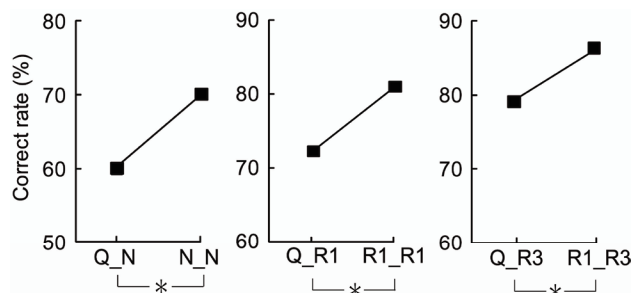


Fig. 7 Results of listening tests for the approach 2 in each listening condition (left, noise; middle and right, reverberation). See Table 2 for the abbreviations on the horizontal axis. [Data is replotted from the reference 23]

響ではマスキングのメカニズムは異なるが、雑音下に加え、残響下の発話も静かな環境での発話よりも有意に明瞭になった。また、発話と聴取が同じ残響環境 (R1\_R1) のみならず、異なる場合 (R1\_R3) においても、残響下の方が静かな環境よりも正解率が有意に上昇した。

以上より、残響音を聞きながら発話することにより、音声案内をより明瞭にすることができる可能性が示された。さらに、その際に使用する残響条件は、音声案内を拡声する公共空間と必ずしも同じである必要はないことも示された。

## 4. おわりに

公共空間における音声案内の明瞭度を改善するため、人の音声情報処理に基づく信号処理と音声生成による手法を提案した。信号処理による手法では、音声の定常部を予め抑圧してから拡声することで、残響によるマスキングを軽減する。音声生成による手法では、人の発話の特徴を活かして、音響的特徴が強調されるように発話環境の設定を行う。雑音・残響下における聴取実験の結果、両手法ともに音声明瞭度の有意な改善が得られた。

今後の課題としては、音声によるバリアフリー化に向けて、公共空間での音声聴取により近い環境での提案手法の評価を行い、信号処理の拡声システムの導入、明瞭な音声案内の録音や合成環境の設定等を行いたい。また、公共空間での音声伝達分野においてもバリアフリー・ユニバーサルデザインのガイドライン化を目指したい。

## 5. 謝辞

本研究は文部科学省科学研究費補助金 (基盤研究 A, 16203041; 若手研究 B, 21700203), 日本学術振興会特別研究員奨励費 (DC2), 文部科学省私立大学学術研究高度化推進事業上智大学オープンリサーチセンター (人間情報科学プロジェクト), 三井住友海上福祉財団研究助成 (高齢者福祉部門) の助成を受けて行われた。

インパルス応答を提供して下さった橘秀樹先生, 上野佳奈子先生, 横山栄先生, 録音と聴取実験に参加して下さいました方々に感謝する。

## 参考文献

- 1) P. D. Danes and E. N. Pinson, "The speech chain: The physics and biology of spoken language," Worth Publishers, 1993 (Second Edition).
- 2) A. K. Nablek and P. K. Robinson, "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am., 71, 1242-1248, 1982.
- 3) 内閣府, "平成22年度版高齢社会白書," 2010.
- 4) 進藤美津子, "老人性難聴の自覚意識の評価法の開発とリハビリテーションに関する研究," 科学研究費補助金 (基盤研究C2) 研究結果報告書, 2002.



- 5) 国土交通省, “高齢者, 障害者等の移動等の円滑化の促進に関する法律 (バリアフリー新法),” 2006.
- 6) 国土交通省, “公共交通機関の旅客施設に関する移動等円滑化整備ガイドライン (バリアフリー整備ガイドライン (旅客施設編)),” 2007改訂.
- 7) 佐藤洋, “公共の音案内の標準化に向けて: 現地調査,” 音サインシンポジウム資料, 2010.
- 8) 総務省, “消防法,” 2003 (改正).
- 9) P. Krishnamoorthy and S. R. Mahadeva Prasanna, “Temporal and spectral processing methods for processing degraded speech: A review,” IETE Technical Review, 26(2), 137-148, 2009.
- 10) K. Kinoshita, M. Delcroix, T. Nakatani and M. Miyoshi, “Suppression of late reverberation effect on speech signal using long-term multiple-step linear prediction,” IEEE Trans. Audio Speech Lang. Proc., 17(4), 534-545, 2009.
- 11) S. Furui, “On the role of spectral transition for speech perception,” J. Acoust. Soc. Am., 80(4), 1016-1025, 1986.
- 12) A. K. Nabelek, T. R. Letowski and F. M. Tucker, “Reverberant overlap- and self-masking in consonant identification.” J. Acoust. Soc. Am., 86, 1259-1265, 1989.
- 13) T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto and T. Kitamura, “Effects on suppressing steady-state portions of speech on intelligibility in reverberant environments,” Acoust. Sci. Tech., 23(4), 229-232, 2002.
- 14) N. Hodoshima, T. Goto, N. Ohata, T. Inoue and T. Arai, “The effect of pre-processing approach for improving speech intelligibility in a hall: Comparison between diotic and dichotic listening conditions.” Acoust. Sci. Tech., 26(2), 212-214, 2005.
- 15) N. Hodoshima, T. Arai, A. Kusumoto and K. Kinoshita, “Improving syllable identification by a preprocessing method reducing overlap-masking in reverberant environments,” J. Acoust. Soc. Am., 119(6), 4055-4064, 2006.
- 16) N. Hodoshima, W. Yoshida and T. Arai, “Improving consonant identification in noise and reverberation by steady-state suppression as a preprocessing approach.” Proc. Interspeech, 1793-1796, 2008.
- 17) N. Hodoshima, P. Svensson and T. Arai “Comparison of consonant identification improvements by steady-state suppression via a loudspeaker system between with and without natural sounds from a talker in reverberation.” Acoust. Sci. Tech., 30(1), 59-62, 2009.
- 18) T. Arai, N. Hodoshima and K. Yasu, “Using steady-state suppression to improve speech intelligibility in reverberant environments for elderly listeners.” IEEE Trans. Audio, Speech, Lang. Proc., 18(7), 1775-1780, 2010.
- 19) W. Van Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow and M. A. Stokes, “Effects of noise on speech production: Acoustics and perceptual analysis,” J. Acoust. Soc. Am., 84, 917-928, 1988.
- 20) J. C. Junqua, “The Lombard reflex and its role on human listeners and automatic speech recognizers,” J. Acoust. Soc. Am., 93, 510-524, 1993.
- 21) 程島奈緒, 荒井隆行, 栗栖清浩, “雑音・残響下における発話の音響的特徴の話者変動,” 電子情報通信学会技術報告, 109(308), 43-48, 2009,
- 22) T. Arai and N. Hodoshima, “Enhanced speech yielding higher intelligibility for all listeners and environments.” Proc. Interspeech, 1620-1623, 2010.
- 23) N. Hodoshima, T. Arai and K. Kurisu, “Intelligibility of speech spoken in noise and reverberation.” Proc. ICA, 2010.