# 拡声音に対する零挿入による残響環境下での音声明瞭度改善の試み －挿入する位置と時間長の検討－*

☆川島佑亮，荒井隆行，安啓一（上智大・理工）

## 1　はじめに

残響環境下において，音声明瞭性が低下する原因の一つとして，overlap-masking [1]という現象が挙げられる．この現象は，直前の音素に付加された残響の尾が，後続の音素に重なることで，音声の明瞭性を低下させるものである．

残響環境下における明瞭性改善には，発話速度を遅くすることが有効である[2-3]．しかし，発話速度を遅くすることにより，音声が伸長されてしまい，エネルギが大きいにもかかわらず音声知覚に比較的重要ではない定常部分の残響の尾が，音声知覚に重要な子音部分や遷移部をマスクしてしまう[4]．従って，overlap-masking の観点からみると，ただ単に発話速度を遅くすることよりも overlap-masking 量を減らすことが重要である．そこで，荒井[5]，松風 [6]らは，それぞれの音節を時間軸上で切り離すことが有効だと考え，定常部の中央に 50, 100 ms の長さの零系列を挿入した．その結果として，単音節における明瞭性の改善が確認されている．しかし，どこの箇所にどれだけの時間長で，零系列を挿入することが最適であるかは，その詳細がまだわかっていない．そこで，本研究では，残響環境下において単語了解度試験を行い，明瞭性が改善する零挿入位置と時間長を調査するとともに，発話速度を遅くした音声とも明瞭性を比較する．

## 2　実験

### 2.1 原音声

NTT-AT 親密度別単語了解度試験用音声データベース(FW03) [7]より，単語親密度 5.5～4.0 の日本語 4 モーラの単語を 26 語用いた．そして，音声合成ソフトウェア「SpeeCAN SFT5（アルカディア社製）」の男性話者を用いて，キャリアフレーズ「これから流す単語は」と，ターゲットである 4 モーラ語を各々作成し，それらを連結することによって，「これから流す単語は○○○○」（○○○○ : 4 モーラ語からなるターゲット語）という刺激文を作成した．なお，キャリアフレーズは 1 つだけ作成し，すべてのターゲットに対し同一のキャリアフレーズを用いるようにした．次に，この音声に対し音声分析・合成ソフト Praat [8]を使用し PSOLA (Pitch Synchronous Overlap and Add)法[9]により，ピッチを変えずに発話速度を 7 mora/s としたものを作成し，これを原音声とした．その際，キャリアフレーズとターゲットのつながりにおいて聴感上の不自然さが生じないよう，キャリアフレーズとターゲットの境界は常に 30 ms の無音になるようにした．

### 2.2 零挿入処理

零挿入処理[5,6]は，ある長さ（*Tz*）の零系列を音声区間に挿入するもので，音素と音素を時間軸上で切り離す手法である．これにより，音素に畳み込まれた残響が，挿入された零系列の区間で減衰することで，後続する音素への overlap-masking 量を減らす効果が得られる．本研究において，零挿入位置は，以下のように，(a)モーラ間，(b)文節間とした．

(a) モーラ間挿入
　こ/れ/か/ら/な/が/す/た/ん/ご/は/○/○/○/○

(b) 文節間挿入
　これから/ながす/たんごは/○○○○

ここで「/」部分に零挿入を行った．

零挿入時間長 *Tz* は，表 1 の通りである．まず，(a)モーラ間挿入における *Tz* を表 1 のように定めた．そして，*Tz* を挿入した際の音声全体の時間長を求めた（表 2）．さらに，この音声全体の時間長がそろうように，(b)文節間挿入における *Tz* を表 1 のように定めた．結局，L1 から L4 のそれぞれにおいては，(a)モーラ間挿入と(b)文節間挿入で音声全体の時間長

* Improving speech intelligibility in reverberant environments using a zero padding technique as a preprocess: Changing length of zeros inserted between moras or phrases, by KAWASHIMA, Yusuke, ARAI, Takayuki and YASU, Keiichi (Sophia Univ.)

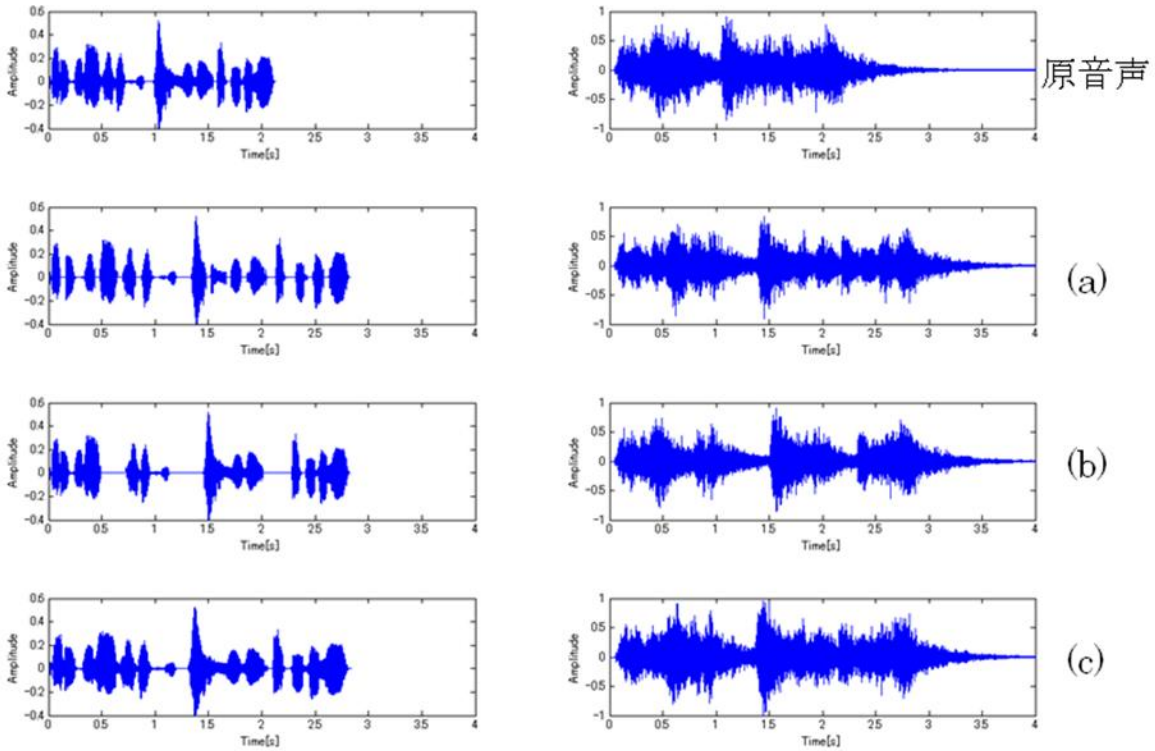


図 1　零挿入処理による音声信号の時間波形（ターゲット語は「あとがま」，L2 の場合）
左図：原音声と処理音声，右図：原音声と処理音声に残響を付加
(a)モーラ間挿入，(b)文節間挿入，(c)全体伸長

は等しく，また L1 から L4 へと添え字の番号が大きくなるに従って音声全体の時間長が長くなる．

表 1　零挿入時間長 $Tz$

| | 零挿入時間長$T_z$ (ms) | | | |
|---|---|---|---|---|
| | L1 | L2 | L3 | L4 |
| (a) モーラ間 | 10 | 50 | 100 | 150 |
| (b) 文節間 | 46.7 | 233.3 | 466.7 | 700.0 |

表 2　L1～L4 における音声全体の時間長

| | 原音声 | L1 | L2 | L3 | L4 |
|---|---|---|---|---|---|
| 全体の時間長(s) | 2.14 | 2.28 | 2.84 | 3.54 | 4.24 |

### 2.3 全体伸長

音声を PSOLA 法[9]で伸長することにより，発話速度を下げた刺激文も比較のため作成した．表 2 の音声全体の時間長に揃うように，発話速度（1 秒あたりの平均モーラ数）を定めた．結果として得られた文の発話速度の平均を表 3 に示す．

表 3　発話速度

| | 平均速度 (mora/s) | | | |
|---|---|---|---|---|
| | L1 | L2 | L3 | L4 |
| (c) 全体伸長 | 6.6 | 5.3 | 4.2 | 3.5 |

### 2.4 使用単語の選択

原音声の単語了解度が 100%であるか確かめるために，合成音声の単語了解度の影響を調べた．データベース内の単語親密度 5.5～4.0 の日本語 4 モーラ語を 56 語用いて，実験で使用する発話速度である 7 mora/s に変換後，単語了解度試験を行った．実験参加者 3 名による結果を表 4 に示す．表 4 より，得られた合成音声はほぼ 100%の了解度であることがわかる．そして，実験で使用するための 26 単語を，了解度が 100%であるという条件で 56 単語の中から選んだ．

表 4　予備実験の単語了解度

| | 実験参加者 | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 正解率[%] | 98.2 | 100.0 | 98.2 |

### 2.5 刺激音

実験刺激は，原音声と，原音声のモーラ間に零挿入を施した(a)_L1～L4，文節間に零挿入を施した(b)_L1～L4，さらに，全体を伸長した(c)_L1～L4 のそれぞれにインパルス応答（残響時間 RT=2.1 s，東京国際フォーラムで測定されたもの）を畳み込んだ音声とした．ゆえに，計 13 条件である．単語 26 語に対し

て全 13 条件で処理を施したので, 総刺激数は計 338 刺激である.

### 2.6 実験手順

実験は防音室内でコンピュータを用いて行った．刺激は外付けオーディオインターフェース(EDIROL UA-25EX)に接続したヘッドホン(SONY MDR-CD900ST)から提示した．実験中は刺激音を一度だけ提示した後に，聞こえたターゲットを仮名でキーボードから入力させた．実験参加者一人につき， 26 語それぞれに異なる 13 条件を割り振り，同じ単語が 2 度提示されないように，計 26 刺激提示した．そして，条件と単語との組み合わせが参加者ごとに異なるようにカウンターバランスをとった．また，刺激の順番はランダムに提示した．

## 3 結果

実験参加者 39 名による, 各条件に対する正解率の平均値を図 2 に示す．ターゲットの 4 モーラ全てが正解した単語数をカウントし，正解率を求めた．横軸は条件を示し，(a)~(c)はそれぞれ, (a):モーラ間挿入, (b):文節間挿入, (c):全体伸長を表す. また, L1～L4 は表 1, 3 に示した零挿入時間・発話速度に対応する．なお，rev とは残響を付加したことを意味する．最も正解率が高かったのは，(b)_L4_rev で 71.8%であった． 次いで， (a)_L2_rev，(c)_L3_rev ともに 70.5%であった.

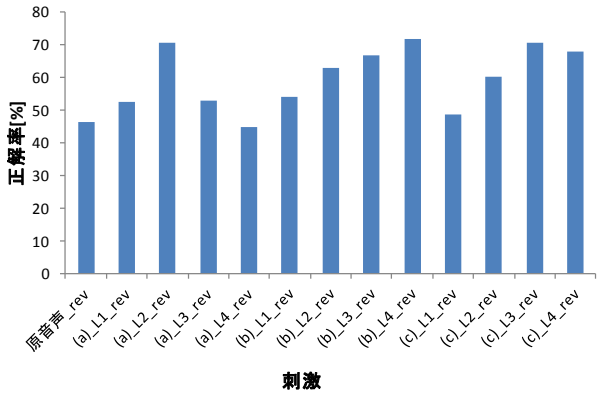


図 2　各条件に対する正解率

統計分析ソフトウェア SPSS を用いて Tukey の多重比較を行ったところ, 原音声_rev と(a)_L2_rev，原音声_rev と(b)_L3_rev，原音声_rev と(b)_L4_rev, 原音声_rev と(c)_L3_rev, 原音声_rev と(c)_L4_rev， (a)_L4_rev と (b)_L4_rev ， (a)_L4_rev と (c)_L3_rev ， (a)_L4_rev と(c)_L4_rev の間で有意差($p < 0.05$)が見られた. 従って, RT = 2.1 s において, (a)モーラ間挿入では $Tz$ = 50 ms，(b)文節間挿入では $Tz$ = 466.7 ms と 700.0 ms において，有意に単語了解度が改善された．また，全体伸長では，発話速度 3.5 mora/s と 4.2 mora/s において有意に単語了解度が改善された.

## 4 考察

### 4.1 零挿入処理について

原音声と比較して有意差が見られたのは, (a)_L2_rev, (b)_L3_rev, (b)_L4_rev, (c)_L3_rev, (c)_L4_rev の条件である．従って，モーラ間挿入，文節間挿入いずれも有意に単語了解度が改善した．その中でも，文節間挿入による改善の効果は大きい.

このように，残響環境下における零挿入処理の効果が確認された．これは，音素と音素の間に零系列を挿入することにより，overlap-masking 量を減らすことができたと考えられる．しかし，モーラ間挿入と文節間挿入で単語了解度を比較すると，モーラ間に $Tz$ =50 ms の零挿入を施したときを除いて，文節間挿入の正解率が大きく高い．さらに，$Tz$ =150 ms でのモーラ間挿入の正解率は，原音声より低い．従って，零挿入を行う位置と零挿入時間長 $Tz$ により, 零挿入処理の効果が異なる.

### 4.2 全体の音声時間長を定めた場合の了解度比較

零挿入処理を実環境に実装する場合，あらかじめある一定の長さに定められた音声全体の時間長の中で，零挿入処理を施さなければならないことがある．その際，音声全体の時間長が同じものの中から，了解度が高い処理を選ぶべきである．そのため，音声全体の時間長別に，それぞれの処理の単語了解度について検討する．表 5 は，各 L1～L4 ごとにおけるそれぞれの処理方法の正解率順位を表している．この表から，零挿入する時間長や伸長時間によって，適切な処理方法が異なる．しかし，文節間伸長は，どの L1～L4 においても, 正解率が比較的上位に入る傾向がある．従って，モーラ間挿入，文節間挿入，全体挿

入を比較して，文節間伸長が全般的に了解度が高いと考える．

表 5　L1～L4 における処理別正解率順位

| | 処理方法 | | | |
|---|---|---|---|---|
| | 高← | 正解率 | | →低 |
| L1 | (b) | (a) | (c) | (d) |
| L2 | (a) | (b) | (c) | (d) |
| L3 | (c) | (b) | (a) | (d) |
| L4 | (b) | (c) | (d) | (a) |

（(a)：モーラ間挿入，(b)：文節間挿入，(c)：全体伸長，(d)：原音声）

### 4.3 ターゲット前の無音区間による影響

表 6 には，モーラ間挿入，文節間挿入，全体伸長におけるターゲット前の無音区間の長さを示した．ここで，L3 の条件での全体伸長と文節間挿入での単語了解度とターゲット前の無音区間の長さに注目する．図 2 より，L3 での全体伸長と文節間挿入の単語了解度はそれぞれ 70.5%，66.7%と差はそれほどない．一方で，表 6 より，無音区間の長さは，それぞれ 47.4 ms，496.7 ms と大きく異なる．この両者を比べて，正解率がさほど変わらないにもかかわらず，ターゲット前の無音区間長は大きな違いがある．この理由としては全体伸長は音声全体を伸長しているので，音声の無音部分と発話部分の両者が伸長されている．そのため，ターゲット前の無音区間が伸長されるという効果と発話部分を遅くしたことによる効果の両方が，了解度に影響したと考えられる．従って，明瞭性の高い音声とは，発話部分を聞き取りやすい発話速度にすること，また，overlap-masking 量を減少させるために，ターゲット前に適度な無音区間を挿入することが明瞭性を改善できるためのポイントであると考えられる．

表 6　ターゲット前における無音区間長

| | 無音区間 (ms) | | | |
|---|---|---|---|---|
| | L1 | L2 | L3 | L4 |
| (a)モーラ間挿入 | 40.0 | 80.0 | 130.0 | 180.0 |
| (b)文節間挿入 | 76.7 | 263.3 | 496.7 | 730.0 |
| (c)全体伸長 | 31.7 | 38.7 | 47.4 | 56.1 |

## 5　おわりに

本研究では，残響環境下において，音声明瞭性が改善をする零挿入位置と時間長 *Tz* を調査した．実験の結果から，文節間に 466.7 ms と 700.0 ms，モーラ間に 50 ms，零挿入をした処理において，有意に単語了解度が改善することが確認された．そして，文節間挿入では，各 L1～L4 の発話長で，比較的高い正解率を示している．従って，文節間に零挿入することによって，全般的に単語了解度が改善されるものだと考えられる．

今回の実験では，モーラ間と文節間に零挿入を行ったが，先行研究[5,6]では定常部に零挿入している．今後は，モーラ間・文節間だけではなく，定常部への零挿入や，異なる残響時間での明瞭性も評価していきたい．


## 謝辞

実験で使用したインパルス応答を提供して下さった東京大学生産技術研究所（当時）の橘秀樹先生，上野佳奈子先生，横山栄先生，そして今回，実験に参加して下った参加者の皆様に感謝申し上げます．

本研究は，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「人間情報科学研究プロジェクト」の支援を受けて行われた．


## 参考文献


[1] Nabelek, *et al*., *J. Accoust. Soc. Am*., 86(4), 1259-1265, 1989.

[2] 翁長他, 日本建築学会計画系論文集, 520, 17-23, 1999.

[3] 四釜他，日本建築学会近畿支部研究報告書, 1-4, 2006.

[4] Arai, *et al*., *Acoust. Sci. Tech*., 28(4), 282-285, 2007.

[5] Arai, *et al*., *Acoust. Sci. Tech*., 26(5), 459-461, 2005.

[6] Matsukaze, *et al*., *Acoust. Sci. Tech*., 2012.

[7] 天野他，親密度別単語了解度試験音表 (FW03)，2003.

[8] Boersma, Praat:doing phonetics by computer [Computer program], Version 5.2.37, retrieved from http://www.praat.org(2011).

[9] Charpentier, *et al*., *Proc. ICASSP*., 2015-2018, 1986.